Dissertationen aus der
Philosophischen Fakultät der Universität des Saarlandes

Slavic Receptive Multilingualism: Intercomprehension of Speech

Jacek Kudera

universaar
Universitätsverlag des Saarlandes
Saarland University Press
Presses Universitaires de la Sarre

OLMS

Jacek Kudera

Slavic Receptive Multilingualism: Intercomprehension of Speech

universaar
Universitätsverlag des Saarlandes
Saarland University Press
Presses Universitaires de la Sarre

OLMS

Universitätsverlag des Saarlandes
Saarland University Press
Presses Universitaires de la Sarre
Postfach 151141, 66041 Saarbrücken

D 291

Dissertation vom 04.05.2022
Dissertation zur Erlangung des akademischen Grades eines Doktors der Philosophie der Philosophischen Fakultät der Universität des Saarlandes
Dekan: Univ.-Prof. Dr. Augustin Speyer
Berichterstatter: Prof. Dr. Bernd Möbius, Prof. Dr. Bistra Andreeva

ISBN: 978-3-86223-316-8 gedruckte Ausgabe
ISBN: 978-3-86223-317-5 Onlineausgabe

Satz: Jacek Kudera
Umschlaggestaltung: Julian Wichert

Bibliografische Informationen der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet abrufbar unter: https://portal.dnb.de/opac.htm abrufbar.

Gedruckt auf FSC-zertifiziertem Papier.

Eine Publikation in Zusammenarbeit zwischen *universaar* und dem Georg Olms Verlag.

Acknowledgements

I would like to thank my mentor, Prof. Dr. Bernd Möbius, for the guidance of my work in the Phonetics Lab at Saarland University. I also wish to express my sincere words of gratitude to all members of the Incomslav project for creating a stimulating interdisciplinary environment and maintaining a good spirit.

None of this would have been possible without my teachers from the University of Wrocław, Prof. Dr. Włodzimierz Wysoczański, Dr. Miłosz Bukwalt, and other passionate lecturers from the Institute of Slavic Studies. Special gratitude to Prof. Dr. William McGregor, who supervised my linguistic studies at Aarhus University and to Prof. Dr. Stefan Werner from the University of Eastern Finland for mentorship in Joensuu.

Kindest words of thanks to Prof. Dr. Vesna Mildner, and Prof. Dr. Vlasta Erdeljac from Zagreb University for leading me through the field of experimental linguistics. Special thanks to Dr. Jürgen Trouvain for the inspiring discussions on paremiology and the morning walks at Prinzenweiher.

This effort would be meaningless without my family to whom I dedicate this work.

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation). Project 232722074: SFB 1102 Information Density and Linguistic Encoding.

Abstract

This work attempts to describe the phenomenon of receptive multilingualism among speakers of the Slavic languages. This dissertation presents a series of one production and four perception studies conducted on a group of native speakers of four Slavic languages, i.e., Bulgarian, Czech, Polish, and Russian. The proposed studies aim to present evidence for mutual comprehension on various levels of processing of closely related languages. The collection of studies involves multiple levels of linguistic analysis. It starts at segmental and suprasegmental phonetic level, then expands the scope to online processing of lexical units and sentences and arrives at cognitively complex tasks of translation of figurative expressions in auditory modality. The applied methodology, involving information theory and experimental phonetics allows for quantification of the cross-lingual intelligibility phenomena among speakers of Slavic languages. The dissertation presents the results of the experiments such as: LADO (language analyses for the determination of origin), EMA (electromagnetic articulography), cross-lingual lexical decision task in short-term priming framework, eye tracking visual world paradigm study, as well as open set and closed set translation test of idiomatic phrases. The results of the production and perception studies were analyzed using quantitative methods. The translation experiment involved both quantitative and qualitative approaches.

The presented studies show that spoken intercomprehension among users of the Slavic languages is asymmetric. It appears that mutual intelligibility varies depending on levels of language analyses. In line with previous investigations on foreign language comprehension, high task-dependency effect was confirmed. The conducted experiments provide evidence for the fact that mutual understanding is often driven by stimuli (un)expectedness. Such an effect appears to be stronger than surface phonetic resemblance of perceived stimuli and corresponding unit

in subjects' native lexicon. Furthermore, the perception studies show that mutual intelligibility does not correspond with the typological divisions of the Slavic languages. The results show that native speakers of Slavic languages can exhibit a privilege in understanding spoken stimuli coming from a language which does not belong to the same grouping. The applied balanced set of tasks involving online and offline language processing, as well as testing material coming from different languages of the Slavic family, led to conclusions which shed light on the multidimensional character of Slavic receptive multilingualism.

The parts of this dissertation concerning the perception and production experiments have already been published or are under review by the time of submission. The details are provided at the beginning of each chapter of this work, which is organized as follows.

The theoretical part of this dissertation introduces previous studies focusing on receptive multilingualism and outlines the methodological apparatus of information theory, which serves as a framework for quantification of phenomena of mutual intelligibility among speakers of closely related languages. The theoretical passages relate the foundations of spoken intercomprehension to regularities in diachronic sound changes and phonological development of the Slavic languages. Since this work is concerned with spoken modality of intercomprehension, the second section of the theoretical part presents a short comparative description of Slavic historical phonology. The empirical part of the dissertation is organized in a structuralist spirit. It starts from the smallest distinguishable units of a language and their role in the auditory identification of linguistic origins by speakers of four Slavic languages. This part is followed by a pilot study involving electromagnetic articulography. This investigation attempts to discover the influence of the information-theoretic notion of surprisal, measured in bits, on highly constrained units of speech such as consonant-to-vowel transitions in acoustic and kinematic domains. Next, the work expands the scope to mutual intelligibility of lexical entries and investigates the influence of associative and phonetic priming on latency times in lexical decision task. It further describes the intercomprehension on a sentence level using an eye tracking study in a visual world paradigm. The last section of the experimental part concerns a complex translation task and cross-lingual understanding of the phraseological expressions in the auditory modality. Such an approach attempts to describe the receptive multilingualism focusing on a particular system of a language. It aims to discover the correspondences and to a certain extent also the asymmetries in Slavic intercomprehension on various levels of language processing. These five studies, which involved 629 native speakers of the Slavic languages, were conducted to understand and describe spoken receptive multilingualism among users of closely related languages. Each study constitutes a separate chapter of the dissertation.

Chapter 1 provides the insights into former studies of receptive multilingualism. It refers to a number of previous investigations in this field and defines the linguistic features which contribute to and impede cross-lingual understanding of speech. This section also defines a new approach in the research on spoken intercomprehension which arose from the combination of Slavic studies, information theory and experimental phonetics. This part introduces the methodological apparatus used in this study and provides and outline of the methods of quantification of receptive multilingualism adapted from the information theory.

Chapter 2 briefly describes the typology of the Slavic languages. It presents the divisions of the Slavic languages in subgroups. This chapter lists the features which were driven by a parallel development of the Slavic languages and may potentially influence spoken intercomprehension. This section also describes the Slavic microlanguages, ethnolects and minority languages. This part draws a picture of genetically and typologically coherent group and points to a regular character of sound changes across the Slavic subgroups. The theoretical foundation for the investigation of Slavic receptive multilingualism is laid down by outlining the diachronic development of the Slavic branch from the Indo-European construct to contemporary languages. Particular emphasis was given to the common linguistic features but also to the divergence processes identified as potential obstacles to mutual intelligibility. Such features were then contrasted with the rest of the members of the Slavic family.

Chapter 3 opens the empirical part of this thesis. It examines the segmental and suprasegmental features of the Slavic languages and their influence on auditory identification of linguistic origin. In the first experimental passages, the study concentrates on the common phoneme inventories of the tested languages. The third chapter presents the results of an auditory language of origin identification experiment in which stimuli involving four Slavic languages were presented to L1 speakers of Bulgarian, Czech, Polish, and Russian with no linguistic training. The goals of the test were to verify the ability of lay listeners to recognize the linguistic origin of speakers, based on spoken samples with limited segmental and suprasegmental information, and to correlate the signal features with the subjects' performance. On the suprasegmental level, it was found that word stress distribution is not an important predictor in the recognition of a language of origin. However, inherent vowel characteristics, such as duration and vowel space computed by the means of Pillai scores correlate with subjects' performance. Both the linguistic profile and the familiarity with closely related languages also appear to be relevant predictors of listeners' performance. Finally, the information-theoretic notion of surprisal applied on regular cross-linguistic sound correspondences was correlated with lay listeners' results. It was concluded that auditory identification of linguistic origin by native speakers of closely related languages is possible even when they are exposed to only limited information such as vowel quality, which can serve as a cue in the identification of linguistic origin.

Chapter 4 is concerned with the relation between the information-theoretic notion of surprisal and articulatory gestures in Polish consonant-to-vowel transitions. It addresses the question of the influence of diphone predictability on spectral and kinematic trajectories by relating the effect of surprisal with motor fluency. The study combines the computation of locus equations with kinematic data obtained from an electromagnetic articulograph. Three groups of Polish diphones were distinguished based on their surprisal values. The kinematic and acoustic data showed that a small coarticulation effect was present in the high- and low-surprisal clusters. Despite of a number of small discrepancies across the measures, a high degree of overlap of adjacent segments is reported for the mid-surprisal group in both domains. This effect is due to both of the following explanations. The first refers to low-surprisal coarticulation resistance and suggests the need to disambiguate predictable sequences. The second, observed in high surprisal clusters, refers to the prominence given to emphasize the unexpected concatenation. These results are consistent with previous studies into the relation between coarticulation strength and contextual predictability. This pilot study provides evidence of the effect of diphone surprisal on the atriculatory gesture. It also argues for an inherent character of surprisal effects in language production.

Chapter 5 presents the outcomes of a lexical decision task in priming technique. It reveals the relation between surprisal, phonetic distance, and latency based on a multilingual, short-term priming framework. Four Slavic languages (Bulgarian, Czech, Polish, and Russian) were investigated across two priming conditions: associative and phonetic priming, involving true cognates and near-homophones. This research proposes new methods for quantifying differences between meaningful lexical

primes and targets for closely related languages. It also outlines the influence of phonetic distance between cognate and non-cognate pairs of primes and targets on response times in a cross-lingual lexical decision task. The experimental results show that phonetic distance moderates response times only in Polish and Czech, whereas the surprisal-based correspondence effect is an accurate predictor of latency for all tested languages. The information-theoretic approach of quantifying feature-based alternations between Slavic cognates and near-homophones appears to be a valid method for latency moderation in the auditory modality. The outcomes of this study suggest that the surprisal-based (un)expectedness of spoken stimuli is an accurate predictor of human performance in multilingual lexical decision tasks.

Chapter 6 presents the results of the tracking visual world paradigm study. It concerns real-time comprehension of spoken stimuli coming from a non-native language tested in a visual environment. In this experiment, a scalable webcam eye-tracker Webgazer was used for gaze estimation and the detection of visual field preferences. Auditory stimuli consisted of a fixed SVO-type sentences in four Slavic languages. The intervals from the verb offset to the onset of a filler, and from the filler offset to the object onset, were equalized across the utterances to ensure an equal temporal scope of the fixation analysis. All the verbs were transitive and controlled for high collocation strength with their respective direct object. The participants were instructed to listen to the sentences and watch the pictures. Head movements were unrestricted during the recording session. The experiment setup precluded the use of a chin rest, but the head pose was tracked in the background. The times to the first fixation on a target visual field with a picture of a direct object were correlated with phonetic distances and surprisal values. Data concerning fixations on the visual target before the verbal component of the audio was played were discarded from the analysis. Such fixations were treated as random and not triggered by information carried by the sentence predicate, which had not yet been perceived. The collected data support the hypothesis that sentence processing in a closely related, non-native language is driven by the information-theoretic notion of surprisal measured on corresponding predicates. Furthermore, the data exhibit the asymmetrical character of intercomprehension across the four groups of Slavic native speakers. This study also supports an argument for surprisal-driven intelligibility effect among speakers of closely related languages.

Chapter 7 presents the results of a translation experiment involving phraseological expressions in Polish and Russian. In order to test cross-lingual understanding of idiomatic units, a series of experiments in the auditory modality involving native speakers of Polish and Russian was conducted. For this study, an information-theoretic approach was adopted by interpreting the experimental outcomes in terms of mean phrase adaptation surprisal, lexical distance, and so–called normalized InDel (insertion/deletion) metrics. The phrase-adaptation surprisal metric was computed by averaging the length-normalized phoneme-adaptation surprisal values of the aligned word pair equivalents in each language. The phrase adaptation surprisal pertains to probabilities of phoneme correspondences of the aligned phrases in Polish and Russian and moderates the understanding of phraseological expressions. The lexical distance measure was calculated on the basis of the proportions of cognates, partial cognates, non-cognates, and false friends among the aligned word pairs of the two phrases. The normalized InDel is a measure of syntactic distance calculated as the total number of inserted or deleted words in a phrase or sentence alignment, divided by the total length of the alignment. In the translation experiment the performance of participants, who were exposed to auditory stimuli in non-native language was tested. The experimental session consisted of two parts. First, in an open set free translation task, the subjects were instructed to directly type their answers. Then, in the subsequent closed set task, participants were asked to select the best match from three predefined variants that represented (1) a semantic equivalent, taken as the entry from a phraseological dictionary; (2) a lemma-based equivalent, sharing a cognate component but differing in the rest of the phrase; and (3) a literal translation of the source, highly divergent from the target idiomaticity. The results of the free translation test were analyzed as a comparison between a target phrase and the provided answer by calculating alterations observed on the construction level, as well as by quantifying differences in the placement of phrase-aligned constituents. Different strategies were discovered across the two translation tasks. The effect of strong surface phonetic similarities of phrases seems to motivate the equivalent matching, especially with respect to literal translation equivalents. Cognate lemma-based identification also plays a role in naive translation, with lexical distance threshold often serving as an idiomatic key. Overall, the data suggest that phonetic, lexical, and syntactic measures of idiomatic phrase pairs can provide an explanation for strategies used by native

speakers of closely related languages in the selection of phrasal correspondences. The results also point out the asymmetries in translations provided by different groups of native speakers. The outcomes of this study contribute to an understanding of the importance of phonetic, lexical and syntactic cues in the process of equivalent selection for isolated idiomatic phrases.

The final part of the dissertation summarizes the experimental findings. It presents the mutual intelligibility of users of the Slavic languages with regard to asymmetries of language pairs and addresses the task dependencies on the intelligibility effects. This section draws the conclusions on receptive multilingualism in the Slavic area and points out the linguistic factors which both contribute to and decrease spoken intercomprehension among the users of closely related languages.

Zusammenfassung

In der vorliegenden Arbeit wird der Versuch unternommen, das Phänomen der rezeptiven Mehrsprachigkeit unter den Sprechern slawischer Sprachen zu beschreiben. Diese Dissertation präsentiert eine Reihe von einer Produktions- und vier Wahrnehmungsstudien, die an einer Gruppe von Muttersprachlern folgender slawischer Sprachen durchgeführt wurden, d. h. Bulgarisch, Tschechisch, Polnisch und Russisch. Die vorgestellten Studien beabsichtigen den Beweis für eine gegenseitige Komprehension auf verschiedenen Sprachbearbeitungsebenen nahverwandter Sprachen zu liefern. Diese Studienreihe umfasst verschiedene Ebenen einer linguistischen Analyse. Sie beginnt mit der segmentalen und suprasegmentalen Phonetikebene, erweitert dann den Skopus auf die Online-Bearbeitung von lexikalischen Einheiten und Sätzen, bis hin zu kognitivkomplexen Translationsaufgaben von figurativen Äußerungen auf der akustischen Ebene. Die angewandte Methodologie, darunter Informationstheorie und Experimentalphonetik, ermöglichte die Quantifikation von sprachenübergreifenden Verständlichkeitsphänomenen unter Sprechern slawischer Sprachen. In der Dissertation werden die Ergebnisse folgender Untersuchungen vorgestellt: LADO (Sprachanalyse zur Bestimmung der Herkunft), EMA (Elektromagnetische Artikulographie), sprachenübergreifende, lexikalische Entscheidungsfragen in einem kurzfristigen Primig-Bezugsrahmen, Eye-Tracking-Studien in Rahmen eines Visual World Paradigms, sowie offene und geschlossene Translationstests idiomatischer Phrasen. Die Ergebnisse der Produktions- und Wahrnehmungsstudien wurden mithilfe quantitativer Methoden ausgewertet. Bei der Auswertung der Translationsaufgabe wurden sowohl quantitative als auch qualitative Herangehensweisen eingesetzt.

Die vorgestellten Studien zeigen, dass die Interkomprehension des Gesprochenen unter den Sprechern slawischer Sprachen asymmetrisch

ist. Es erscheint, dass die gegenseitige Verständlichkeit abhängig von der Sprachanalyseebene variiert. Wie in vorherigen Untersuchungen der fremdsprachlichen Komprehension wurde auch in diesem Fall eine hohe Aufgabendependenz festgestellt. Die durchgeführten Studien liefern einen Beweis dafür, dass gegenseitige Verständlichkeit oft von der (Un)Vorhersehbarkeit der Stimuli abhängig ist. Dieser Effekt erschient stärker als oberflächliche phonetische Ähnlichkeit der wahrgenommen Stimuli oder korrespondierende Einheiten im Lexikon der Muttersprache von den Probanden zu wirken. Darüber hinaus zeigen die Wahrnehmungsstudien, dass die gegenseitige Verständlichkeit nicht mit der topologischen Unterteilung der slawischen Sprachen übereinstimmt. Die Ergebnisse beweisen, dass Muttersprachler slawischer Sprachen ein Privileg beim Verstehen von gesprochenen Stimuli, die aus einer nicht zu derselben Gruppierung gehörenden Sprache kommen, aufweisen können. Das angewandte ausgewogene Set mit Online- und Offline-Sprachbearbeitungsaufgaben, sowie das Testmaterial, das aus verschiedenen Sprachen der slawischen Sprachfamilie stammt, führt zu Schlussfolgerungen, die ein neues Licht auf den multidimensionalen Charakter der slawischen rezeptiven Mehrsprachigkeit werfen.

Die Dissertationsteile, die die Produktions- und Wahrnehmungsstudien beinhalten, wurden bereits veröffentlicht oder befinden sich zur Zeit der Dissertationseinreichung in Begutachtung. Genauere Angaben werden zu Beginn jedes Kapitels der Arbeit gemacht, die folgend strukturiert wurde.

Der theoretische Teil der vorliegenden Dissertation beinhaltet die Einführung in bereits existierende Untersuchungen der rezeptiven Mehrsprachigkeit und eine Beschreibung der Instrumentarien der Informationstheorie, was als Bezugsrahmen für die quantitative Untersuchung der gegenseitigen Verständlichkeit unter Sprechern nahverwandter Sprachen dienen soll. Weiter wird auch der Zusammenhang zwischen den Grundlagen der Interkomprehension des Gesprochenen und den Regelmäßigkeiten in den diachronischen Lautveränderungen und phonetischen Entwicklungen in den slawischen Sprachen angesprochen. Da sich diese Arbeit auf die Interkomprehension des Gesprochenen konzentriert, wird im weiteren Abschnitt des theoretischen Teils kurz auf die komparative Phonologiegeschichte der slawischen Sprachen eingegangen. Der empirische Teil der Dissertation wird in dem strukturalistischen Geist gehalten. Ausgegangen wird von den kleinsten unterscheidbaren Spracheinheiten und deren Rolle in der akustischen Identifikation der Herkunftssprache von Sprechern vier unterschiedlicher slawischer

Sprachen. Weiter stellt man eine Pilotstudie vor, die sich der elektromagnetischen Artikulographie bedient. Es hat zum Zweck, den Einfluss des informationstheorie-basierten Surprisalwertes, der in Bits gemessen wird, auf stark eingeschränkte Spracheinheiten wie Konsonant-Vokal-Übergänge im akustischen und kinetischen Bereich herauszufinden. Im weiteren Verlauf breitet man den Anwendungsbereich auf die gegenseitige Verständlichkeit der lexikalischen Einheiten aus und untersucht den Einfluss von assoziativen und phonetischen Priming auf die Latenzzeit, in der die Entscheidung bezüglich der Lexikwahl getroffen wird. Als nächstes wird die Interkomprehension auf der Satzebene verwendet, indem man Eye-Tracking-Studien in Rahmen eines Visual World Paradigms einsetzt. Der letzte Absatz des empirischen Teils stell eine komplexe Translationsaufgabe und die zwischensprachliche Verständlichkeit der Phraseologismen auf der akustischen Ebene vor. Dank solcher Herangehensweise erscheint es möglich, die rezeptive Mehrsprachigkeit in Bezug auf ein bestimmtes Sprachsystem zu beschreiben. Man beabsichtigt auch die Ähnlichkeiten und in bestimmten Umfang auch die Unterschiede in der slawischen Interkomprehension auf verschiedenen Ebenen der Sprachenbearbeitung zu beschreiben. Diese fünf Studien, in deren Rahmen insgesamt 629 Sprecher von slawischen Sprachen untersucht wurden, sollen dazu beitragen, die rezeptive Mehrsprachigkeit des Gesprochenen unter Sprechern nahverwandter Sprachen zu verstehen und zu beschreiben. Jeder Studie wurde ein getrennter Kapitel dieser Dissertation gewidmet.

Kapitel 1 liefert den Einblick in bereits existierende Untersuchungen der rezeptiven Mehrsprachigkeit. In diesem Kapitel wird auf die vorherigen Forschungen auf diesem Feld eingegangen. Zusätzlich werden auch sprachliche Merkmale definiert, die zu einer zwischensprachlichen Verständlichkeit des Gesprochenen beitragen oder sie hemmen. In diesem Teil findet sich ebenfalls die Erläuterung einer neuen Herangehensweise im wissenschaftlichen Umgang mit der Interkomprehension des Gesprochenen, der seinen Ursprung in der Slawistik, Informationstheorie und Experimentalphonetik hat. Weiter wird das methodologische Instrumentarium der Studie vorgestellt und der Bezugsrahmen von den Quantifizierungsansätzen der rezeptivem Mehrsprachigkeit geschildert, die der Informationstheorie entnommen und angepasst wurden.

In Kapitel 2 findet sich eine allgemeine Typologie der slawischen Sprachen. Präsentiert wird die Unterteilung der Sprachen in Untergruppen. Hier werden auch die Merkmale aufgelistet, die während einer simultanen Entwicklung der slawischen Sprachen entstanden und einen Einfluss auf die Interkomprehension des Gesprochenen haben können. Dieser Teil beinhaltet ebenfalls eine Beschreibung der slawischen Mikrosprachen, Ethnolekte und Sprachen der Minderheiten. Dieser Teil schildert weiter eine genetisch und typologisch kohärente Gruppe und weist auf einen regemäßigen Charakter der Lautveränderungen unter den Untergruppen der slawischen Sprachen hin. Die Basis zur Erforschung der rezeptiven Mehrsprachigkeit der slawischen Sprachen wird durch die diachronische Entwicklung des slawischen Zweiges aus dem indoeuropäischen Stamm zu den zeitgenössischen Sprachen unterstützt. Eine besondere Aufmerksamkeit wird den allgemeinen Sprachmerkmalen aber auch den Divergenzprozessen geschenkt, die als mögliche Hürden bei der gegenseitigen Verständlichkeit erscheinen können. Diese Merkmale werden folglich mit anderen Vertretern der slawischen Sprachfamilie verglichen.

Kapitel 3 eröffnet den empirischen Teil der Dissertation. Hier werden die segmentalen und suprasegmentalen Merkmale der slawischen Sprachen und deren Einfluss auf die akustische Identifikation der Herkunftssprache untersucht. In den ersten Fragmenten des Experiments konzentriert man sich auf das allgemeine Phoneminventar der untersuchten Sprachen. Es werden auch die Ergebnisse des Experiments zur akustischen Identifikation der Herkunftssprache gezeigt, in dem L1-Sprecher folgender vier slawischer Sprachen: Bulgarisch, Tschechisch, Polnisch und Russisch, mit Sprachstimuli konfrontiert wurden. Zu betonen ist, das die Probanden weder eine sprachliche Ausbildung genossen haben noch Sprachkenntnisse der angegebenen Sprachen aufwiesen. Das Ziel dieser Untersuchung war es festzustellen, ob die Laienzuhörer in der Lage sind, die Herkunftssprache der Sprecher zu erkennen, während sie Aussagen mit begrenzten segmentalen und suprasegmentalen Informationen hören, und Signalmerkmale mit der Leistung der Probanden zu verbinden. Auf der suprasegmentalen Ebene hat man herausgefunden, dass die Verteilung der Wortbetonnung keinen bedeutenden Einfluss auf die Erkennung der Herkunftssprache hat. Jedoch, inhärente Vokalmerkmale, solche wie Länge oder Raum berechnet nach den Pillai-Werten, stehen in Wechselbeziehung mit der Leistung der Probanden. Es erscheint jedoch, dass sowohl das Sprachprofil als auch die Vertraulichkeit mit nahverwandten Sprachen wichtige Faktoren bei der Leistung der Zuhörer sind. Zuletzt wurde auch bewiesen, dass der informationstheoretische Surprisalwert, der auf regemäßige zwischensprachliche Laute übertragen wurde, in einer Wechselbeziehung mit der Leistung der Laienzuhörer steht. Daraus kann die Schlussfolgerung gezogen werden, dass eine akustische Identifizierung der Herkunftssprache bei Muttersprachlern nahverwandter Sprachen möglich ist, auch wenn ihnen nur begrenzte Informationen, wie z. B. Vokalqualität, zur Verfügung gestellt werden, was als ein Hinweis in der akustischen Identifizierung der Herkunftssprache dienen kann.

In Kapitel 4 wird die Beziehung zwischen dem informationstheoretischen Surprisalwert und den artikulatorischen Gesten in den Konsonant-Vokal-Übergängen des Polnischen erforscht. Es wird auch die Frage nach dem Einfluss der Diphonen-Vorhersehbarkeit auf die spektralen und kinematischen Verläufe in Bezug auf den Surprisaleffekt und die Sprachbeherrschungsmotorik angesprochen. Diese Studie vereint die Berechnung der Ortsgleichung mit den kinematischen Angaben, die aus einem elektromagnetischen Artikulographen gewonnen wurden. Anhand des Surprisalwertes wurden drei Gruppen von polnischen Diphonen ausgesondert. Die kinematischen und akustischen Daten haben gezeigt, dass ein kleiner Koartikulationseffekt bei den Surprisalbündeln mit hohem und niedrigem Wert entstand. Obwohl es kleine Wiedersprüche bei den Bemessungen gab, konnte eine Überlappung der angrenzenden Segmente in dem Surprisalbündel mit mittlerem Wert in beiden Bereichen festgestellt werden. Dies kann zweierlei gedeutet werden. Einerseits bezieht sich das auf den Wiederstand bei Koartikulation des Surprisalbündel mit niedrigem Wert und schlägt das Klärungsbedürfnis bei vorhersehbaren Sequenzen vor. Andererseits, was bei Surprisalbündeln mit hohem Wert beobachtet wird, deutet es auf den Bedarf, eine unerwartete Verknüpfung in den Vordergrund zu stellen. Diese Ergebnisse bestätigen die bereits existierenden Studien im Bereich der Beziehungen zwischen der Koartikulationsstärke und der kontextuellen Vorhersehbarkeit. Diese Pilotstudie liefert die Beweise für den Einfluss des Diphonen-Surprisaleffekts auf die artikulatorischen Gesten und plädiert für den inhärenten Charakter des Surprisaleffekts auf die Sprachenproduktion.

In Kapitel 5 werden die Ergebnisse einer lexikalischen Entscheidungsaufgabe in der Priming-Methode vorgestellt. Basierend auf einem mehrsprachigen, kurzfristigen Primig-Bezugsrahmen zeigt man die Verhältnisse zwischen der Surprisalfunktion, phonetischer Distanz und Latenzzeit. Vier slawische Sprachen (Bulgarisch, Tschechisch, Polnisch und Russisch) wurden in Bezug auf zwei Priming-Voraussetzungen untersucht: den assoziativen und phonetischen Priming mit Berücksichtigung von realen Kognaten und Fast-Homophonen. In Rahmen dieser Untersuchung wird eine neue Methode der Quantifizierung der Unterschiede zwischen bedeutsamen lexikalischen Primes und Zielen bei nahverwandten Sprachen vorgestellt. Es hebt auch den Einfluss der phonetischen Distanz zwischen kognaten und nicht kognaten Wortpaaren auf die Entscheidungszeit bei einer zwischensprachlichen lexikalischen Aufgabe hervor. Die Ergebnisse zeigen, dass die phonetische Distanz nur im Polnischen und Tschechischen einen Einfluss auf die Entscheidungszeit hat, während der auf der Surprisalfunktion basierende Korrespondenzeffekt ein genauer Hinweis auf die Latenzzeit in allen vier untersuchten Sprachen ist. Die informationstheoretische Herangehensweise bezüglich der Quantifizierung der merkmalbasierten Veränderungen zwischen Kognaten und Fast-Homophonen in den slawischen Sprachen erscheint als eine berechtigte Methode, um die Latenzunterschiede auf der akustischen Ebene zu erforschen. Die Ergebnisse deuten auch darauf hin, dass die auf dem Surprisaleffekt beruhende (Un)Vorhersehbarkeit der gesprochenen Stimuli ein angemessener Hinweis auf die menschliche Leistung bei mehrsprachigen lexikalischen Entscheidungsaufgaben ist.

Der Kapitel 6 stellt die Ergebnisse des Eye-Tracking-Experiments in Rahmen des Visual World Paradigms vor. Es wurde das Echtzeitverständnis der akustischen Stimuli, die nicht aus der eignen Muttersprache stammen, in einer visuellen Umgebung getestet. Während des Experiments wurde die skalierbare Eye-Tracking-Kamera Webgazer benutzt, um die Blickveränderungen einzuschätzen und Präferenzen beim visuellen Feld aufzudecken. Die akustischen Stimuli bestanden aus fixen SVO-Satztypen in vier slawischen Sprachen. Die Zeitspanne zwischen dem Verb-Offset und dem Filler-Onset, und auch von dem Filler-Offset zum Objekt-Onset wurde unter allen Aussagen ausgeglichen, um einen gleichmäßigen Zeitumfang der Fixationsanalyse zu gewehrleisten. Alle verwendeten Verben sind transitiv und weisen eine sehr starke Kollokation mit dem angegebenen Objekt auf. Die Probanden wurden dazu aufgefordert, den Aussagen zuzuhören und sich die Bilder anzuschauen. Kopfbewegungen waren während der Aufnahme nicht eingeschränkt. Die Experimentausstattung schloss die Verwendung einer Kinnstütze aus, aber die Kopfstellung wurde im Hintergrund mitverfolgt. Die Zeit bis zu ersten Fixation auf dem visuellen Feld mit einem Bild des Objekts wurde mit der phonetischen Distanz und den Suprisalwerten in Verbindung gebracht. Daten, die die Fixation auf ein visuelles Objekt vor dem Abspielen des Audiokomponenten beinhaltetet, wurden aus der Analyse ausgeschlossen. Solche Fixationen wurden als zufällig und nicht durch Informationen aus den Stimulisätzen ausgelöst betrachtet, da die Informationen noch nicht verarbeitet werden konnten. Die gesammelten Daten belegen die Hypothese, dass die Verarbeitung der Sätze aus einer nahverwandten Nichtmuttersprache durch die informationstheoretische Suprisalfunktion, die anhand korrespondierender Prädikate gemessen wird, geleitet werden. Darüber hinaus veranschaulichen die Daten einen asymmetrischen Charakter der Interkomprehension in allen vier Gruppen der slawischen Muttersprachlern. Dies unterstütz auch das Argument für den surprisal-basierten Verständlichkeitseffekt unter den Sprechern nahverwandter Sprachen.

In Kapitel 7 werden die Ergebnisse eines Translationsexperiments mit phraseologischen Äußerungen zwischen Polnisch und Russisch in beide Richtungen geschildert. Um die zwischensprachliche Verständlichkeit der idiomatischen Wendungen zu untersuchen, wurde eine Serie von akustischen Experimenten mit polnischen und russischen Muttersprachlern durchgeführt. Für diese Studie wurde die informationstheoretische Herangehensweise angepasst, indem man die erhaltenen Befunde in Rahmen des Phrasenadaptationssurprisals, der lexikalischen Distanz und des so genannten normalisierten InDels (insertion/deletion) analysiert hat. Um die Werte für den Phrasenadaptationssurprisal auszurechnen, hat man sich der Durchschnittsangaben der normalisierten Länge des Phonemadapationssurprisalwertes bei den angepassten Wortpaaräquivalenten in der jeweiligen Sprache bedient. Der Phrasenadaptationssurprisal bezieht sich auf die Wahrscheinlichkeit der Phonemkorrespondenz in den zugeordneten Phrasen im Polnischen und Russischen und moderiert die Verständlichkeit phraseologischer Ausdrücke. Die lexikalische Distanz wurde anhand der Häufigkeit der Kognaten, partieller Kognaten, Nicht-Kognaten und „falscher Freunde“ unter den verwendeten Wortpaaren in den zwei Phrasen ausgerechnet. Das normalisierte InDel veranschaulicht die syntaktische Distanz, die mithilfe der Anzahl von ausgelassenen oder hinzugefügten Wörtern in der Phrase oder im Satz, die dann durch die Länge der ganzen Gruppierung geteilt wird, ausgerechnet wird. In dem Translationsexperiment wurde die Leistung der Probanden gemessen, die akustischen Stimuli in der Nicht-Muttersprache ausgesetzt wurden. Das Experiment bestand aus zwei Teilen. Erst wurden die Probanden in einer offenen Translationsfrage gebeten, ihre Antworten gleich zu tippen. Zweitens wurden die Probanden in einer Multiple-Choice-Frage gebeten, aus den drei angegebenen Antwortmöglichkeiten die ihrer Ansicht nach Beste zu wählen. Darunter fanden sich folgende Möglichkeiten: (1) semantisches Äquivalent, das aus einem Phraseologie-Wörterbuch entnommen wurde; (2) ein lemma-basiertes Äquivalent, das einen Kognaten teile aber Unterschiede im restlichen Teil aufwies; (3) eine wortwörtliche Übersetzung des Ausgangstextes, die große Unterschiede von der idiomatischen Redewendung in der Zielsprache aufwies. Die Ergebnisse des ersten Tests wurden im Hinblick auf den Vergleich zwischen der Zielsprachenphrase und der gelieferten Antwort analysiert, indem man sowohl die beobachteten Änderungen auf der Konstruktionsebene ausgerechnet, als auch die Unterschiede in der Stellung der Phrasenkonstituenten quantifiziert hat. Zusätzlich hat man auch unterschiedliche Strategien bei der Bewältigung der zwei Translationsaufgaben entdeckt. Es erscheint, dass starke oberflächliche phonetische Ähnlichkeiten in Phrasen die Auswahl motivieren, insbesondre mit Berücksichtigung des wörtlichen Äquivalents. Kognate lemma-basierte Identifikation spielt auch eine Rolle in der Laientranslation, wobei die lexikalische Distanzgrenze oft als ein idiomatischer Schlüssel dient. Im Allgemeinen zeigen die gewonnenen Daten, dass phonetische, lexikalische und syntaktische Werte von idiomatischen Phrasenpaaren eine Erklärung der von Muttersprachlern nahverwandter Sprachen benutzten Strategien bei der Wahl von Phrasenkorrespondenz liefern können. Die Ergebnisse weisen ebenfalls Ungleichheiten in Übersetzungen auf, die von unterschiedlichen Muttersprachlerngruppen geliefert wurden. Die Befunde dieser Studie tragen auch dazu bei, dass die Wichtigkeit der phonetischen, lexikalischen und syntaktischen Hinweise im Prozess der Äquivalentauswahl bei isolierten idiomatischen Phrasen anerkannt wird.

Der letzte Teil dieser Dissertation beinhaltet die Zusammenfassung aller Experimentergebnisse. Es wird die gegenseitige Verständlichkeit der slawischen Spechern mit Berücksichtigung der Unterschiede zwischen den Sprachenpaaren präsentiert und die Aufgabenabhängigkeit auf den Verständlichkeitseffekt angesprochen. In diesem Teil finden sich auch die Schlussfolgerungen über die rezeptive Mehrsprachigkeit unter den slawischen Sprachen. Darüber hinaus werden sprachliche Faktoren angegeben, die zu der Interkomprehension des Gesprochenen unter Sprechern nahverwandter Sprachen beitragen und sie auch hemmen.

Contents

Chapter 1

Introduction

1.1 Aims and rationale

This work attempts to investigate phenomena concerning mutual intelligibility of speech among native speakers of the Slavic languages. It aims to present evidence for inter-language intelligibility on various levels of non-native language processing. It addresses a niche in a scientific intercomprehension discourse by applying a hybrid methodological apparatus that combines experimental phonetics and information theory. Furthermore, this work aims to shed light on linguistic features that may facilitate or inhibit intercomprehension. Since a degree of intelligibility varies across modalities, present work narrows the focus to speech understanding exclusively. Presented collection of studies involves multiple levels of linguistic analyses. It starts from phonetic segmental and suprasegmetal perspective, then shifts the attention to articulatory phonetics, moves on to processing of the cognate and non-cognate lexical entries, extends the scope to sentence comprehension and finishes with complex task of cross-lingual understanding of idiomatic phrases. Such a narrative permits an investigation of receptive multilingualism for several systems of language, while having in mind the effects of task-dependencies as well as structural properties of tested languages. Also, the methodology of conducted experiments can reflect the differences between online and offline language processing. The applied balanced set of tasks enables to draw the conclusions that address the multi-directional character of comprehension. Finally, a multilingual perspective involving a set of four

Slavic languages (Bulgarian, Czech, Polish, and Russian) may point to differences and asymmetries in speech comprehension.

Investigating receptive multilingualism is motivated not only by presenting the comprehensible features of closely related languages and deliberations on their origins. The fields of areal linguistics and dialectology can certainly benefit from empirical studies on lingua receptiva and enrich the findings by relating particular isoglosses to linguistic areas of higher and lower mutual intelligibility. Apart from purely linguistic interest, such studies can be vital to educational policy makers who, when in possession of large sample studies on receptive multilingualism, may efficiently design the education curricula. Moreover, education of second language learners, as well as training of translators, who work with closely related languages can be supplemented with reports on high and low intelligibility structures across languages, which may result in more accurate equivalent selections. Needless to say that studies on inherent intelligibility can serve as empirical evidence for language reconstruction and diachronic comparative descriptions of language varieties. As shown in this work, another applicable aspect of intelligibility studies concerns the LADO (language analysis for the determination of origin) procedures, especially in cases of limited reference data. Finally, a cultural dimension of intercomprehension studies should be stressed. Studies on receptive multilingualism entail the aspects of understanding across the groups of users of languages grounded not only on related lexicons, but also coming from a similar cultural heritage.

1.2 Receptive multilingualism

Receptive multilingualism refers to a form of communication in which speakers use their own native languages, rather than learnt L2s, due to a high degree of mutual intelligibility of their languages. Genetic relatedness of used varieties, which often results in their typological proximity, can drive efficient and relatively fluent cross-lingual conversation. Several motivations can trigger multilingual interactions. For some, an ease of expression in a native language can motivate using L1. Also, one may choose to use genetically close languages to express certain cultural concepts shared by both varieties, but rather hard to describe by means of a lingua franca. Apart from the question of linguistic competence, some may choose to stick to their own native language to mark a linguistic identity and emphasize divergent behaviour. However, an ideal cross-lingual manner of interaction presupposes fairness and equality between parties which make an effort to understand an interlocutor without showing any advantage in mastering a third language. Such a communication practice in the literature has been called *semicommunication.* This term, coined by (Haugen 1966) will not be used in this work due to its inherent attribute, which implies an incomplete or partial communication. Quantification of this kind seems to be an accurate description of only selected communicative strategies. The interchangeable terms such as *receptive multilingualism*, *intercomprehension*, *lingua receptiva*, *inherited multilingualism*, *nonconvergent discourse*, *asymmetric/bilingual discourse*, *cross-language communication*, and *plurilingual communication* also refer to the above-mentioned interaction mode and are widely applicable in the scientific discourse (Gooskens 2013; Rehbein et al. 2012; Casad 1974; Gooskens and Heuven 2021; Thije and Zeevaert 2007). A terminological extension of receptive multilingualism was proposed by (Verschik 2012). A specification of communication mode in which sides use their own L1s was called *inherent receptive multilingualism*, opposed to so called *acquired receptive multilingualism*, which presupposes a certain degree of knowledge of an interlocutor's language. According to this dichotomy, the empirical part of this work is concerned with the inherent receptive multilingualism investigated by the means of functional tests on a pool of speakers of closely related languages of the Slavic family.

Previously, a degree of mutual intelligibility has been considered as a criterion in investigation of genealogical relationship between languages (H. Hickerson and N. Hickerson 1952). Interdialectal intelligibility has been suggested as a measure of dialect distance (Voegelin and Harris 1951) and further interpreted by (Wolff 1959), as an indicator of cultural and societal relationships. The pioneering studies into intercomprehension (Voegelin and Harris 1951) introduced a difference between *reciprocal* and *non-reciprocal* intelligibility indicating possible asymmetries in comprehensions. Furthermore, a distinction between *mutual intelligibility* and *neighbor intelligibility* has been proposed. This dichotomy has been founded on the premises that the former results from a genetic relatedness, whereas the latter emerges from language contact. A number of previous studies on receptive multilingualism in the European context have regarded Scandinavian communication practices (Thije and Zeevaert 2007; Braunmüller 2007; Gooskens 2007; Doetjes 2007; Haugen 1966) as well as Slavic (Stenger 2019; Jágrová et al. 2019; Bulatović et al.

2019; Sloboda and Nábělková 2013; Stenger and Avgustinova 2021; Golubović 2016), Germanic (Bezooijen and Gooskens 2007; F. H. E. Swarte 2016; Gooskens and Heuven 2017; Beerkens 2010), and Romance intercomprehension (Jensen 1989; Melo-Pfeifer 2014; Donato and Pasquarelli-Gascon 2015; Bonvino 2015). The older studies into Slavic intercomprehension focused on calculation of similar lexical entries and proposed the percentages of very similar words as indicators of mutual understanding (Sławski 1962). In line with the intelligibility scoring, several attempts have been made to estimate a comprehension-based threshold for dialect vs. language distinction (Casad 1974), though, such calculations seem to have an arbitrary character.

A historical overview of the European research projects focusing on intercomprehension was given by (Bonvino 2015). Apart from obvious linguistic relatedness, a number of research initiatives have examined factors which influence mutual intelligibility, such as speakers' age and level of education (Ploeg et al. 2017; Vanhove 2014) or personality traits (Lambelet and Mauron 2017), and concluded that communicating in a mutually intelligible variety can be motivated not only by a lack of command of any other language but also by individual preferences. This question, related to the Scandinavian area has been raised by (Braunmüller 2008). Several other studies have already shown that important factors contributing to mutual intelligibility cover prejudices, social stigmas and attitudes towards members of certain speech communities and their languages (Schüppert et al. 2015; Gooskens and Bezooijen 2006; Tang and Heuven 2007). Despite the relevance of attitude testing, in this work, only functional tests, not opinion scores have been undertaken. According to the typology of intelligibility tests (Gooskens and Heuven 2017; Gooskens 2013), this work refers to the methods examining functional and perceived intelligibility (Tang and Heuven 2007) rather than judged comprehension. The overview of the intelligibility tests applied to different language families, as well as assets and disadvantages of the testing methods, was given by (Gooskens 2013). Differences between receptive and productive multilingualism have been summarized by (Braunmüller 2007). They mainly refer to communicative situational context and style. In the comparison mentioned above, a pragmatic dimension of an interaction has been stressed. Receptive multilingualism has been assigned to informal purpose-oriented communication circumstances and face to face oral interaction gravitating towards situational contexts. On the contrary, a function-oriented communication type along with dominance of linguistic awareness of styles and norms have been mentioned as characteristics of productive multilingualism.

Perception studies on dialect intelligibility have been driven by various motivations. A pragmatic dimension of such initiatives aimed to establish a common orthography for the need of developing the literacy programs (Casad 1974). A degree of mutual understanding has been applied as a canonical criterion in distinguishing between languages and dialects (Voegelin and Harris 1951; Wolff 1959). In the following work, however, such a dichotomy will be a subject to challenge because speakers of certain language areas within the Slavic family exhibit lower inter-dialectal intelligibility when compared to other languages. This motivates rather unrestricted use of terms *languages* and *vernaculars* in the scope of this work. Since non-L1 language processing is measured by a set of modality-dependent tests, perception studies on receptive multilingualism have implemented a distinction between stimulus modalities. Intelligibility of text certainly varies from speech understanding, not only due to various scripts, but also due to differences in orthography based on similar alphabets, such as different functions of diacritics across languages (Stenger 2019). In short, opaque orthography, which does not clearly correspond to spoken language, causes more difficulties in intercomprehension than transparent systems with more direct grapheme to phoneme relations.

Investigations into receptive multilingualism are vital not only to linguistics, but also to cultural and regional studies. An importance of multilingualism in transnational context finds its expression on the institutional level, as well as in bottom-up initiatives such as *Deklaracija o zajedničkom jeziku* (Declaration on the Common Language) (Bugarski 2018), *Deklaration om nordisk språkpolitik* (Declaration on a Nordic language policy) (Ministerådet 2006), or Final Report High Level Group on Multilingualism (European Communities 2007), which contributes to raising awareness of multilingual communication and education.

1.3 Information theory

Research on intercomprehension so far has been lacking coherent objective methods for predicting intelligibility (Gooskens 2013; Gooskens 2007). A number of previous studies presented an importance of particular linguistic phenomena on intelligibility and concluded that phonetic distance seems to be the most important predictor of intelligibility, followed by word length (Kürschner et al. 2008), which then can be attributed to phonological neighborhood density effect. Previous attempts to establish the importance of single linguistic phenomena in multilingual scenario covered only selective systems of language. A more holistic approach to questions of predictability of lingua receptiva can be inspired by the methodology adapted from information theory. An underling assumption of such an adaptation is based on a premise that listening to speech coming from other than native, but closely related language, resembles the situation in which input is nonoptimal. Therefore, the quantification of such input in relation to presumably optimal signal coming from one's native language can be conducted by implementing the methods of information theory. Hence, by applying a theoretical framework of information theory, this work attempts to examine intercomprehension as a function of contextual predictability.

The foundations of the discipline were laid down by Claude Shannon's *Mathematical Theory of Communication* (Shannon 1948) and then extended by Warren Weaver (Shannon and Weaver 1949; Krippendorff et al. 2009) who established the theory that found the applications far beyond initially intended fields of cryptography and telecommunication. The goal of information-theoretical models of information exchange, firstly proposed to optimize communication between machines, has been defined to maximize the amount of information transferred via a channel, while considering its intrinsic limitations. The most optimal use of the channel capacity, in theoretical premises of the discipline, should result in relatively stable information flow. Such an optimal use of transmission channel avoids peaks and troughs of quantified information across time. A widely applicable character of Shannon's theorems encouraged researchers to treat human communication in a similar way and measure language processing in cognitive tasks as well as computational models (Crocker et al. 2016). In linguistics, this research apparatus has been further extended and used to address the questions of cognitive effort in language processing (Demberg et al. 2012; Delogu et al. 2017). Previous studies have shown that this methodological apparatus can be successfully applied on various levels of language processing such as phonetic segmental and suprasegmental (Shaw and Kawahara 2019; A. Bell et al. 2009; Jurafsky et al. 2001), phonological (Raymond et al. 2006; Hume 2016), morphophonological (Shaw et al. 2014), lexical (Seyfarth 2014), as well as syntactic (Jaeger 2010; Demberg et al. 2012). In multilingual studies, information-theoretical apparatus has been previously applied by (Piantadosi et al. 2011; Stenger 2019; Stenger and Avgustinova 2021). The combination of linguistic and information-theoretic approaches has led to establishing of several prominent theories such as Probabilistic Reduction Hypothesis (Jurafsky et al. 2001), Smooth Signal Redundancy Hypothesis (Aylett and Turk 2004; Aylett and Turk 2006), Uniform Information Density (Jaeger 2010), and Informational Redundancy Hypothesis (Pluymaekers et al. 2005). These theories have been contextualizing the probabilistic effects regarding linguistic encoding and language variability.

Several measures have been proposed to objectively assess the amount of information transferred via a known channel in a given point of time (Hale 2016; Brandt 2019). Metrics, such as surprisal and entropy, have often been used to answer the question of linguistic or acoustic redundancies driven by efficient communication principle and optimal use of channel capacity. Therefore, for the need of this work, an information-theoretic notion of surprisal has been applied to quantify the unexpectedness of input in cross-lingual comprehension and production studies. Surprisal allows for quantification of the (un)expectedness of a unit in its context, and is calculated according to the following equation:

$$S(unit_x) = -log_2 P(unit_x|context)$$

in which *unit* refers to any linguistic element in focus of investigation given its direct *context*. Log transforming of the conditional probability renders the units in bits. From this equation, units that are less predictable score a high surprisal values, and vice versa, contextually predictable elements exhibit low surprisal. In studies examining multilingual correspondences, as well as L1 to L(n) transfer, the notion of surprisal can be applied to test the cross-lingual input adaptation and, therefore, quantify the (un)expectedness of chosen units in exposure to signal coming from non-native language.

Since previous studies regarding information-theoretic concepts have been applied to test the effects of predictability or frequency on linguistic variability, a natural extension of the paradigm could shed light on the effects of information-theoretic notion of surprisal in processing spoken signal coming from a non-native, but closely related language.

Chapter 2

Slavic Languages

The theoretical foundation of receptive multilingualism is built on the structural similarities of closely related languages. The role of similarities across the languages is therefore crucial in successful communication. Furthermore, a relative territorial unity of the area inhabited by the speakers of Slavic languages contributes to mutual intelligibility. The contrastive analysis of features of all investigated languages can help to identify the comprehension cues and potential difficulties in mutual intelligibility. Hence, the following chapter outlines the linguistic features shared within particular subgroups of the Slavic languages and attempts to trace the origins of diverged elements in a diachronic perspective.

2.1 Contrastive diachrony

The branch of the Slavic languages belongs to the satem languages of the Indo-European language family. It is commonly divided into three groups: West, East, and South. This trichotomy was founded on grammatical and lexical similarities of languages which emerged from the brake-up of the Proto-Slavic unity. Even though other conventions of clustering were proposed, nowadays most typologists seem to agree upon the Slavic tripartite.

The historical development of the Slavic languages can be traced back to the Proto-Slavic, and a step further, to the Balto-Slavic period which emerged from the Indo-European construct. Three main periods can be distinguished in the development of the Slavic family. A period of unified language development followed by territorial expansion, which results in dialectal differentiation, and then in consequence leads to a break-up of communicative unity and emergence of different languages entities. Shortly before the time of Proto-Slavic unity, structural similarities between the Slavic and Baltic languages (such as common palatalization of consonants by a close-front vowel, delabialization of labialized velars, similar evolution of prosody as well as corresponding results of apophony) led to the identification of the of Balto-Slavic period. An in-depth analysis of Trautmann's *Baltisch-slavisches Wörterbuch* (Trautmann 1923), also suggests the strong unity of the Balto-Slavic group, even though the Balto-Slavic Hypothesis has been questioned in a vibrant discussion among historical linguistics. Archeology, onomastics, and recently also genetics, trace the first period of the Proto-Slavic to 2000-1500 BC. This period could have been proceeded by the five-century-lasting Balto-Slavic entity (Moszyński 2006; Sławski 1988; Kuryłowicz 1957; Matasović 2005).

The vocalic system of Proto-Slavic must have been derived from the Indo-European inventory which consisted of monophthongs, diphthongs and sonants, that is syllabic consonants. It included closed-front */i/, close back */u/, mid-front */e/, mid-back */o/, open */a/ and */ə/, which in the course of the language development was either reduced or altered into */o/. These monophthongs exhibited the quantitative long vs. short distinction, which remained in symmetrical long vs. short vowel opposition system in Czech and the asymmetric ones in ijekavian Croatian, Slovak and Slovenian (Dalewska-Greń 2007). The system of diphthongs consisted of the following sequences: */eu/, */ou/, */au/, */ei/, */oi/, and */ai/. The syllabic consonants were: */r/, */l/, */m/, and */n/. In the Balto-Slavic period, such systems were prone to the following alternations: merging of the */a/ and */o/ vowels, which in consequence led to symmetry of the system with regard to front vs. back distinction, as well as the emergence of new soft vs. hard opposition of sonants and a gradual process of their division into vocalic and consonantal element. Some of the most frequently mentioned reconstructions on the suprasegmental level refer to a phonemic feature of tone comprised of acute and circumflex. The latter stem-placed intonation was a subject of the Fortunatov–de Saussure's Law, which explains the shift of stress placement within a word, but also alters the stress on the proceeding preposition. These alternations resulted in free stress placement in Russian as well as in the more complex intonation system of Serbian/Croatian. The relics of the Proto-Slavic accentuation system remain in the Čakavian Croatian, which in many cases preserved the stress place before the shift. The reconstructed Proto-Slavic language is still a topic of a linguistic debate and many scholars concerned with the Slavic diachrony presented contradicting theories. However, the features which historical linguists found less disputable cover the open syllable law, which originates in the Law of Rising Sonority as well as the syllabic synharmony.

Changes in the vocalic system, inherited from the Balto-Slavic era, were qualitative and quantitative and concerned the development of close front */*i*/, close-mid front */*ь*/, open-mid front */*e*/, open front */*ě*/, close back */*u*/, close-mid back */*ъ*/, open-mid back */*o*/, and open back */*a*/. As a consequence in Proto-Slavic, the qualitative distinctions took over the quantitative property of short and long segments. In further development the long vs. short distinction was reduced and the close-mid vowels, the so-called *jers*, lowered their articulation, became shorter and either reduced or vocalized depending on their position. Weak *jer* was lost in coda position and before the vowel, whereas the strong *jer* position was held before a syllable with a weak *jer*. The law of open syllables also affected the diphthong clusters comprising vowels and labials and led to their monophthongization, which resulted in the appearance of the nasals. Moreover, the fronting of the elements adjacent to */*j*/ was a process common for all Slavic territory with only a few exceptions (Popowska-Taborska 1984; Sussex and Cubberley 2006; Mareš 1969). Furthermore, the other alternation, the so-called iotation, caused by */*j*/ following the velars resulted in the emergence of the palatal affricates. The first palatalization was caused by the front vowels and resulted in the following changes: */*k*/ > */*k'*/ > */*č*/, */*g*/ > */*g'*/ > */*dž*/ > */*ž*/, and */*x*/ > */*x'*/ > */*š*/. The second regressive palatalization resulted in sibilants /*c*/, /*z*/, and /*s*/. The third palatalization, however, did not occur consequently over the whole Slavic territory. It mainly affected the articulation of velars in codas of nouns.

In the final period of the Common Slavic unity, the metathesis of liquids took place. However, the different reflexes of this process in onsets across the North and South Slavic suggest that the metathesis could have taken place after the migration of the Slavs (Lehr-Spławiński 1957). The differentiation of three groups of the Slavic languages is often association with the great migration of Slavs in the fifth an sixth centuries AD. The directions of migrations from relatively coherent habitat, that is: East towards the Dnieper basin, West reaching the north-east of modern Germany, and South to the Balkan Peninsula and even further to the Peloponnese, are now associated with the main division of the language family. As a result of this migration, the Common Slavic language started its process of gradual divergence, which lasted until the ninth century. In this time, the Great Moravian Mission of Saints Cyril and Methodius initiated the Slavic written culture. The tenth century brings the Common Slavic era to an end. Further alternations of *jers*, and the marginalization of the law of open syllable result in the north vs. south division. The migration era enhanced the differentiation process and fossilized local dialectal alternations, which later became the distinctive features of particular language territories and vernaculars.

In the course of evolution, several Slavic languages became extinct, such as: Old Church Slavonic (South-East Slavic) nowadays only in use in the Orthodox liturgical practice as Church Slavonic, Polabian (West Slavic, Lechitic group), Slovincian (West Slavic, Lechitic group, Pomeranian subgroup) or Knaanic (West Slavic). Regardless of the process of either forced or natural divergence of the languages which constitute the Slavic group, the parallel development and regularities of the alternations within the subgroups still contribute to the strong mutual intercomprehension of speakers of Slavic languages. What is more, the emergence of the transitional stages and dialect continua contributes to intelligibility of the speakers of the Slavic languages and hence lay a foundation for contemporary investigations of the receptive multilingualism of the speakers of the three following groups of the Slavic languages.

2.2 West Slavic

The West Slavic group is canonically divided into three subgroups: Lechitic (also spelled Lekhitic), Czech-Slovak and Sorbian. Five contemporary languages, that is: Polish, Kashubian, Czech, Slovak, Upper Sorbian and Lower Sorbian as well as Polabian, extinct in the eighteenth century, belong to the West Slavic group. The common features of the group are the presence of the alveolar fricative as a result of the second and third palatalization (e.g., Polish *wszech*, Czech *vše*, Lower Sorbian *wšen*, Upper Sorbian *wšón*); the transition of Proto-Slavic **tj*,**kt'* into *c* (e.g., Polish *świeca, noc*, Czech *svíce, noc*); and perservation of the **tl*, **dl* clusters (e.g., Polish *plótł*, Czech *pletl*, Slovak *pletol*, Lower Sorbian *pletł*, Upper Sorbian *pletł*). A more fine-grained division of the group involves the distinction of North subgroup (Polish, Kashubian, Lower and Upper Sorbian) and the Southern subgroup of the West Slavic, namely Czech and Slovak. Such detailed typology is a result of the analysis of the diachronic development in which three clear stages can be distinguished. The first stage common for the whole West Slavic group refers to the qualitative change of back /*ъ*/ and front /*ь*/, lowering the articulation of *jers* and therefore merging into *e*. The further differentiation of West Slavic group led to an interesting quantitative alternation of *jers* which resulted in compensatory lengthening of the proceeding vowel while shortening the *jer*. The weak *jer* in coda became neutralized, whereas the syllabic one in West Slavic was consequently altered into *e*.

The second stage of differentiation, which introduced a North vs. South distinction, resulted in retention of syllabic liquids in the Southern group; whereas the alternation in the Northern territory resulted in vowel + liquid sequences. Moreover, the phonemic vowel length opposition remained in the Southern territory but became neutralized in the Northern part. However, an important exception in the horizontal division can be noted in the regularities of the *d' > dz' > z* shift in Polish, Slovak and *d' > dz' > dz* in Czech and Sorbian. Such development seems to contradict the hotizontal orientation in typology of the West Slavic (Sussex and Cubberley 2006).

The third divergence stage covered the alternations, which resulted in different reflexes in each language of the group. A development of *r'* can be treated as such category (e.g., Polish *rzeka*, Sorbian *rěka* and Slovak *rieka*). From this stage onward, the process of divergence of the members of the West Slavic group does not allow to point out the common features, and historical foundation for the intercomprehension of speakers of the West Slavic languages can be observed on the level of particular subgroups.

2.2.1 Lechitic group

The Lechitic subgroup of the West Slavic languages can be divided into West (extinct Polabian and West-Pomeranian dialects) and East (Polish, Silesian and the Pomeranian subgroup composed of Kashubian and Slovincian - the latter though is sometimes treated as a dialect of Kashubian). Leaving aside sociolinguistic disputes on nomenclature which involves arbitrary terms such as literal, national, regional, dialect, and ethnolect, the division of the Lechitic group, proposed in the 24th edition of Ethnologue covers Polish, Kashubian and Silesian. Since the status of the latter requires separate attention with regard to mutual intelligibility, the several common features for the East Lechitic cluster (Polish and Kashubian) should be summarized. One of the most prominent innovations of the Lechitic subgroup when compared to the West Slavic is a dispalatalization process as well as the perseveration of the nasal vowels (in other Slavic territories present only in peripheral dialects). The process of dispalatalization covered several Proto-Slavic vowels followed by apical hard consonants as well as dispalatalization of primarily syllabic *r* and *l*, even though the traces of such alternation can be found in the Sorbian subgroup as well. The change of *tort* into *tart* can be also described as a subgroup feature with several exceptions in Kashubian, which suggests its transitional character.

The outstanding features of the Polish language, which might present a difficulty in intercomprehension, are the vocalization of ъ into e and elision of the ь. On the suprasegmental level, word stress placement at the penultimate, which evolved from the initial stress distribution typical for the Czech-Slovak and Sorbian subgroups as well as for the South Kashubian, also distinguishes Polish stress distribution pattern. Another feature which distinguishes Polish is the preservation of the nasal vowels *ą* and *ę*. Three of the Polish vernaculars, that is the Lesser Polish, Greater Polish and Mazovian (Urbańczyk 1953) along with their subdialects are mutually comprehensible. The Podhale region holds an exceptional intelligibility status, as it is less comprehensive and in some classifications even treated as a microlanguage. Kashubian (sometimes spelled 'Cassubian') ethnolect is divided into three highly divergent subgropus of Northern, Central and Southern Kashubian. It is characterized by the presence of *tart* group in place of Proto-Slavic **tort*, the alternation of short *i*, *y*, and *u* into 'Kashubian schwa' ë, dispalatalization of /sˈ/, /ʑ/, /ʨ/, /ʥ/ into /s/, /z/ /ʦ/, /ʣ/. (Topolińska 1980). The Silesian ethnolect is also internally diverged. The main phonetic archaism which distinguishes Silesian in its group is the pronunciation of *ą* instead of *ę* (Sławski 1988). Also, in the Silesian vernaculars a high amount of Czech and German loanwords are observed. Silesian therefore seems more difficult to comprehend for Polish standard users. This relation is unilateral, since the majority of Silesian speakers are proficient users of standard Polish.

2.2.2 Czech-Slovak group

A certain distinction of Czech and Slovak when compared with the other members of West Slavic is attributed to the occurrence of syllabic consonants *r* and *l*. These approximants can create syllable nuclei also in the South Slavic group. The development of the *tort*, *tert*, *tolt*, and *telt* clusters also separates the Czech-Slovak group. The phonological distinction of vowel length, which differentiates the Czech-Slovak group, is perceptually important for testing intercomphrenesion in auditory modality. The transition of *št*, *žt* sequences into *št'*, *žd'* (e.g., Proto-Slavic **sъčęstьje* > Czech *štěstí*, Slovak *šťastie*), as well as the alternation of *g* into *h* (e.g, Proto-Slavic **gora* > Czech and Slovak *hora*) diverge the Czech-Slovak group from the other groupings and can potentially influence the intelligibility in interaction with other Slavic native speakers.

The distinctive phenomena present in Czech cover the qualitative alternation of *a* proceeded by a soft consonant into *e*, *ě*, or *ie*. Also, the transition of *u* and *ú* into *i* and *í* after palatals (e.g., Proto-Slavic **tjudjь* > Czech *cizí*) presents an exceptionally Czech alternation. The distinctiveness within a group regards the *jat* / *jat'* vowel which developed differently in Czech and Slovak. In the former, due to the erosion of the iotation process, *jat* turned into *e*, except when positioned after labial consonants, where *je* is preserved; whereas in the latter *ie* sequence remained (e.g., Proto-Slavic **věra* > Czech *víra*, Slovak *viera*). The dyphthongization of long *u* into *ou* (e.g., Proto-Slavic **sušь* > Czech *souš*, Slovak *súš*) is an important feature observed in this group. Also a perpiheral dialectal change of long *y* into *ej* in Czech refers to a similar process but is not noted in standard varieties. Regular alternations of *u* into *ů* (e.g., Proto-Slavic **bogъ* > Czech *bůh*) are also typical for Czech territory. The three main dialectal areas are mutually comprehensible. In the suprasegmental perspective the important Czech vowel lengthening, originating from Balto-Slavic acute intonation, remained in disyllabic lexemes, opposed to Slovak or Polish where acute accent became shortened (e.g., Proto-Slavic **kr̋ova* > Czech *kráva*, Slovak *krava*, Polish *krowa*). Additionally, with regard to length distribution in Slovak, a long vowel cannot be adjacent to another prolonged segment, with an exception in mid-Slovak dialects. Slovak realizations of strong *jer* are *e*, *o*, and *a* (e.g., Proto-Slavic **olkъtь*, **orol*, **pętъkъ* > Slovak *laket'*, *orьlъ*, *piatok*). In Slovak, but not in Czech, the syllabic consonants *r* and *l* can be either long or short, by analogy to the vocalic segments. Also, the *l* sonant in Slovak is syllabic in all its positions. The dialectal landscape of Slovak is rich and can be divided into three groups: Central (with transition of acute and circumflex sequences *or̃t*, *ol̃t* into *rãt*, *lãt*), West (with preserved *št* group), and East (with loss of long vs. short vowel distinction and regular word stress at first syllable, similarly to Polish).

2.2.3 Sorbian group

Several works treated Sorbian subgroup as a transition stage between the other two groups with geographical and typological proximity of Lower Sorbian to Northern subgroup and Upper Sorbian to the Southern one (Sussex and Cubberley 2006). Moreover, the Sorbian subgroup has been treated as a transitional stage between the Lechitic group and Czech. The dialectal differentiation of the Sorbian subgroup is extensive. However, several features are common for both Upper and Lower Sorbian, such as: the retention of dualis, aorist and imperfect. The development of the sequence **tort* into *trot*, as well as the lack of the *tart* cluster are common Sorbian features. Similar to Polish and Kashubian, Sorbian group is characterized by almost unexceptional palatalization of consonants followed by front vowels and preservation of the /ɨ/ vowel. The phonetic characteristics limited only to the Sorbian group cover the simplification of the *str* sequence into the *tr* cluster (e.g., Proto-Slavc **sestra* > Lower Sorbian *sotša*, Upper Sorbian *sotra*). Strong *jer* was vocalized into *e* in Lower Sorbian and into *e* or *o* in the Upper Sorbian (e.g., Proto-Slavic **dъždžь* > Upper and Lower Sorbian *dešć*). The consonantal clusters *tr*, *pr*, and *kr* evolved into Upper Sorbian *tš*, *pš*, and *kš*; Lower Sorbian *tś*, *pś*, and *kś*. Another distinctive phenomenon typical for the Sorbian subgroup is the alternation of *wi* > *ji* > *j* (e.g., Proto-Slavic **praviti* > Upper Sorbian *prajić*).

Even though Upper and Lower Sorbian belong to the same subgroup, their degree of mutual intelligibility is limited. The insular character of these languages led to several differences such as deaffrication of *ć* and *dź* into *ś* and *ź* (e.g., Proto-Slavic **teta* > Lower Soriban *śota*, Upper Sorbian *ćeta*); alternation of *č* into *c* (e.g., Proto-Slavic *časъ*, *čelo* > Lower Sorbian *cas*, *coło*; Upper Sorbian *čas*, *čoło*). Also, different outcomes of the alternations caused by the assimilation of place of articulation are prominent in both languages (e.g., Lower Sorbian *šyja*, *šyś* vs. Upper Sorbian *šija*, *šić*) (Stieber 1965). The qualitative change of /ɑ/ into /ɜ/ when flanked by palatals is another distinctive inner-group characteristic (e.g., Proto-Slavic **aje* > *jaje* Upper Sorbian *jejo*, Lower Sorbian *jajo*). The strong influence of initial word stress in Upper Sorbian results also in the vowel reductions in unstressed positions (e.g., Proto-Slavic *pьšenica* > Upper Sorbian *pšeńca*, Lower Sorbian *pšenica*).

2.3 East Slavic

Three contemporary languages whose speakers show high spoken intercomprehension belong to the East Slavic group: Russian, Belarusian, and Ukrainian. Rusyn minority language also belongs to East Slavic. The East Slavic group is a clear example of high oral intelligibility historical foundation of which can be traced back to thirteenth century AD, when the gradual process of divergence of Old Russian unity began. The transition of *tl*, *dl* > *l* is common for the entire East Slavic territory but also present in the South Slavic languages. Furthermore, the approximants *l* and *r* in the East Slavic territory are not syllabic. On the suprasegmental plane the languages which emerged from the Old Russian eroded the quantitative stress distinction and melodic word accent. Remnants of the acute and circumflex can be observed only in the forms *-ert-*, *-ort-*, *-elt-* which evolved into *-erét-*, *-orót-*, *-olót-* under the acute stress and *-éret-*, *-órot-*, *-ólot-* under circumflex (Sławski 1988). The *tj*, *kt'*, and *dj* development into *č* and *ž* also has a regular character in East Slavic languages (e.g., Proto-Slavic **světja*, **noktь*, **medja* > *свеча*, *ночь*, *межа*). What is more, the presence of epenthetic *l* after the labials characterizes the East Slavic languages (e.g., Proto-Slavic **zemja* > Russian, Ukrainian *земля*; Belarusian *зямля*).

Elision of nasal vowels is another common feature for the whole East Slavic territory. The *ǫ* altered into *u* in the entire East Slavic area, whereas *ę* turned into *a* (e.g., Proto-Slavic **dǫbъ*, **zǫbъ* > Russian *дуб*, *зуб*; Proto-Slavic **pętь*, **tęgnǫti* > Russian *пять*, *тянуть*). The dialectal distinctions which refer to the 3. sg. praes. suffix *-tь* present in peripheral Russian, standard Belarusian, Ukrainian, and North-East Ukrainian dialects and its hard counterpart *-t* in literary Russian, North and Central Russian dialects, and South-West Ukrainian vernaculars do not impact the strong mutual intelligibility among speakers of the East Slavic languages. The onset alternation of *e* into *o* (e.g., Proto-Slavic **ezero* > Russian, Ukrainian *озеро*; Proto-Slavic **elenь* > Russian, Ukrainian *олень*) is a typical feature of the group. However, this shift depending on the stress, resulted in different outcomes in East Slavic vernaculars. This process in stressed position gave *e* > *o* in standard Russian and Belarusian, but in the North Russian language area the shift occurred in unstressed position. In Ukrainian such an alternation did not have a regular character and took place, regardless of the accent after *č*, *ž*, and *š* consonants, but never in coda. The process of vocalization of *jers* (hard into *o* and soft into *e*) did not occur simultaneously across the entire East Slavic area. The elision of soft semi-vowels in North Russian territory took place as late as the second half of the thirteenth century, later than in the South Russian vernaculars.

The qualitative and quantitative reduction of unstressed vowels, mainly caused by the flexible type of lexical stress is typical for the Russian language. In Russian and Belarusian the unstressed *o* and *a* typically turn into short *a* or *schwa* depending on the stress placement. This variation, not reflected in Russian orthography, is often called *akanye*, named after the approximated quality of outcome of vowel reduction. The distinctive features of the Russian language, when compared with the other East Slavic languages cover the stressed Nom. pl. suffix -*a*, not noted in Belarusian and rarely seen in Ukrainian dialects. In Russian, the onset *i*- is preserved as a remnant of Proto-Slavic sequence *jь*- (e.g., Proto-Slavic **jьgrati* > Russian *играть*). A similar alternation can be found in the South Slavic group. However, a development of this sequence in Ukrainian and Belarusian resembles the East Slavic (e.g., Proto-Slavic **jьgrati* > Ukrainian *грати*, Belarusian dialectal *граю*). Similar development concerns the preoposition из / з and из- prefix. Also the Gen. sg. suffixes (-*его*, -*ого*) in pronouns and adjectives distinguishes Russian from Belarusian and Ukrainian. The two former even reflect the difference in more transparent orthography. The other systematic sound change concerned velar stops. In Russian, the Proto-Slavic */*g*/ remained whereas in Ukrainian and Belarusian it changed the manner of articulation, and is consequently realized as a /x/ fricative. Additionally, the C-/j/ clusters and -*ьje* coda type led to gemination of consonants in Ukrainian and Belarusian, but remained ungeminated in Russian (e.g., Russian *веселье*, Belarusian *вяселле*, Ukrainian *весілля*).

Russian dialects are primarily divided into North, South, and Central. The last one exhibits a transitional character between the former two. The North group of Russian vernaculars is divided into five smaller subgroups. This group is characterized among the others by *a* to *e* alternation after the soft consonants, elision of intervocalic *j*, and often assimilation of consonant + nasal clusters into two nasals. From inter-language comprehension perspective, an interesting dialectal characteristics of the North region can be associated with typical South-East Slavic feature. A quasi similarity to Bulgarian postpositive article can be found in the augmentative also postpositive particle (*-ot*, *-ta*, *-to*). The pretonic vowel in the North area is often pronounced as *o* (so-called *okanye*), similarly to the accented vowel. The Southern dialectal territory is more diverged than the Northern part. It is characterized by the regular transitions of unstressed *o* into *a* (*akanye*), merging of *o* and *a* in pretonic syllables, as well as realization of Proto-Slavic *jat* as mid-front or close-mid front vowel. Also, the peripheral character of /*g*/ > /x/ change is noted in the dialects of the South. The transitional Central dialects exhibit *k* and *g* transition into *t'* and *d'* after the front vowels, *akanye* and continuation of velar stop instead of /x/. These characteristics laid a foundation of the contemporary literary Russian.

In Ukrainian, the unstressed /*o*/ pronounced as /*a*/ practically does not exist. Several other features distinguish Ukrainian from the rest of the East Slavic group. These are the development of Proto-Slavic *jat* into /*i*/ (e.g., Proto-Slavic **sněgъ*, **gnězdo* > Ukrainian *сніг*, *гніздо*); a transition of /*e*/ and /*o*/ into /*i*/ (Proto-Slavic **ledъ*, **stolъ* > Ukrainian *лід*, *стіл*); merging of the Proto-Slavic */*i*/ and */*y*/ into one high-mid front vowel; the presence of remaining Proto-Slavic */*c*/ (e.g., *праця*, *місяць*); as well as the absence of final devoicing of coda consonants, also noted in the Serbo-Croatian group. The regressive devoicing in consonantal sequences does not take place, which also applies to the North Russian vernaculars and selected Belarusian dialects. Two principal dialectal areas can be distinguished in Ukrainian, namely, Northern and Southern. One of the isoglosses which constitutes such a division relates to the pronunciation of /*a*/ in stressed and unstressed positions. In the North dialectal area, divided into three subdialects, the /*a*/ segments which do not constitute a syllable nucleus are pronounced as /*e*/ after non-palatalized consonants. In the Southern area, which consists of two sub-regions, the above-mentioned alternation does not happen, regardless of stress distribution. Furthermore, in the Northern part the palatalization of consonants followed by /*i*/ which is a contiunant of long /*o*/, does not take place. A prominent feature of Ukrainian South dialectal area is merging of the vowel mid-front qualities with close-front and central segments. Similarly, mid-back articulation gravitates towards close-back. The Ukrainian literary language was based on the South-East dialectal, considerably unified area.

In Belarusian, *akanye* is more common than in Russian. This alternation (/*e*/, /*o*/ > /*a*/) is triggered by proceeding soft consonants. Furthermore, such a sound change is noted in Belarusian orthography.

The shift of palatalized dental and alveoral stops into soft dental affricates is one of the alternations which distinguishes Belarusian from other East Slavic languages. What is more, the hardening of palatalized */r'/* (e.g., вячэра), in contrast to Russian and partly Ukrainian (e.g., вечеря), differentiates Belarusian from its group. Belarusian dialects are typically divided into two groups, that is, North-Eastern and South-Western. The small area of transitional dialects with selective features of both areas can also be distinguished. In terms of lexis, numerous Polonisms are shared by both Belarusian and Ukrainian, which in addition to regular sound changes, contributes to the effect of mutual intelligibility.

2.4 South Slavic

South Slavic group is divided into two subgroups the Eastern, composed of Bulgarian, Macedonian and Old Church Slavonic; and the Western which consists of Slovenian and Bosnian / Croatian / Montenegrin / Serbian. The diachronic changes typical for the whole South Slavic area cover the development of *ę* into *e*, merging of *y* and *i* and hardening of palatals and dental affricates. The common South Slavic alternations, confirmed on the basis of the Old Church Slavonic texts also regarded the inversion of sequences *ьr*, *ъr*, *ьl*, *ъl* and *rъ*, *rъ*, *lь*, *lъ* (e.g., Old Church Slavonic: *sъmrъtь* > Bulgarian *смърт*, Slovenian, B/C/S *smrt*; Old Church Slavonic *sъnce* > Bulgarian *слънце*, Slovenian *solnce*, B/C/S *sunce*). The two groups, Western and Eastern can also be distinguished on the suprasegmental plane. Not only the stress distribution patterns, but also the word stress type differs across two groups of the South Slavic languages. The combination of pitch and length fluctuation is present only in the Western group. Such a melodic accentuation scheme is not present in the Eastern branch. Another process which diverged the South-West from South-East was linked to the development of the Proto-Slavic **tj*, **kt'*, and **dj* clusters. The degree of aperture in articulation of the vowels which emerged from *jat* also divides the subgroups of the South Slavic languages, namely, in the Bulgarian and Macedonian subgroup the reflexes of *ě* are open, whereas in the South West Slavic languages the sound change resulted in more closed vowels, with exception of closing diphthong in Slovenian vernaculars.

2.4.1 Western group

The Western group of the South Slavic languages comprises Slovenian and highly intelligible Bosnian / Croatian / Montenegrin / Serbian (often abbreviated to B/C/S, B/C/M/S or Serbo-Croatian which adhere to strong intercomprehension effect among speakers of this pluricentric standard). The Western group of the South Slavic languages shares the following features: merging of the ъ, and ь semi-vowels and their vocalization into *a* in B/C/M/S and *a* or *e* in Slovenian (e.g., Proto-Slavic **сънъ* > B/C/M/S *san / сан*, Slovenian *sanja*). The Western group preserved the epenthetic *l* emerged as a consequence of coarticulatory effect of *j* following the labial consonants (e.g., Slovenian and B/C/M/S *zemlja*).

Dialects of the South-West group are based on three realizations of *jat* vowel which developed (**ě* > *e*) in the Ekavian area, (**ě* > *(i)je*) in the Ijekavian territory and (**ě* > *i*) in the Ikavian area. The reflexes of *ě*, and systematic sound changes caused by three possible pronunciations of former *jat* do not influence the cross-dialectal, and naturally, cross-national intelligibility. However, the isoglosses which define the reflexes of *jat* constitute only one part of the Serbo-Croatian dialectal division. The supreme dialectal division is called after the pronunciation of the *what* lexeme and divides the South-West Slavic territory into Štokavian, Kajkavian, and Čakavian dialects called *наречје/narječje* (Ivić 2001). Apart from the differences in interrogative particle, the three dialectal areas are distinguished by the word-final *l* to *o* alternation, differences in palatalization of dental stops caused by adjacent *j*, application of -*ov*- and -*ev*- infixes as plural markers as well as the frequency of aorist verb forms which differs across the areas.

Štokavian standard is shared by all contemporary varieties, that is Bosnian, Croatian, Montenegrin, and Serbian. However, Kajkavian and Čakavian are spoken primarily in modern Croatian linguistic landscape. Kajkavian is spoken in Slovenia, which is being divided into seven dialectal areas of different levels of mutual intelligibility. The standard literary languages of Bosnia, Croatia, Montenegro, and Serbia are considerably more similar than the vernaculars present in modern Croatia (Kordić 2010). Therefore, the basic distinction between standards and dialects should be given exceptional attention in the linguistic landscape of South-West Slavic group. In order to provide detailed classification of particular dialectal area, the combination of description of *jat* reflex and the interrogative particle common for a dialectal region should be provided (e.g., Ekavian-Štokavian, Ikavian-Čakavian etc.). Furthermore, detailed classification of this dialectologically rich area includes a dichotomy between territorially the largest Štokavian area into Neo- or New-Štokavian and Old-Štokavian dialects depending on the development of the accent type. The Serbo-Croatian group is characterized by the four types of lexical stress which includes two variables, namely rising and falling pitch as well as shortening and lengthening of stressed segments. The threefold, archaic type of accent remained in Čakavian (long-falling, short-falling, and long-rising acute). Additionally, in Serbo-Croatian group the vowel lengthening can occur in unstressed position, whereas in Slovenian the vowel lengthening is associated with word stress position. Furthermore, merging of the new acute with long-falling intonation type is an outstanding feature of prosodic landscape in Serbo-Croatian group. The old accentuation type was preserved in Montenegrin and peripheral dialects of Slavonia.

One of the numerous distinctions between Slovenian and Serbo-Croatian regards the vowel reductions. Slovenian short and unaccented vowels are often reduced, as opposed to Serbo-Croatian group. Also, the coda *l* to *o* shift (e.g., Proto-Slavic **koьlъ* > *kotao*) distinguishes the Serbo-Croatian group, with several exceptions noted in vernaculars. The reflexes of the *čr* group, which were preserved in Slovenian but altered into *cr* in Serbo-Croatian group, differentiate the languages of the South-Western group. Slovenian is the only Slavic language in which Proto-Slavic nasal **ǫ* turned into *o* (e.g., Proto-Slavic **zǫbъ*, *pǫtь* > Slovenian *zob*, *pot*; B/C/M/S *zub*, *put*). Additionally, only Slovenian and Lower Sorbian (West Slavic) exhibit the supine verb forms, which in other Slavic languages was merged with infinitives. What is more, the use of dualis in Slovenian declination and conjugation is often mentioned as exceptional in the South Slavic area.

Regardless of the numerous isoglosses dividing the Slavic part of the Balkan peninsula, mutual intercomprehension of the Serbio-Croatian group is not questionable. Even on the lexical level, the analysis of the Swadesh list comprising a hundred so-called basic words, showed that standard national varieties differ in one item only, that is 'a liver' (*džigerica – jetra*), which again is well known by the native speakers of all languages which belong to the South-West Slavic group, so cannot present any obstacle in comprehension (Kordić 2010). Communication between Slovenian native speakers and users of Serbo-Croatian standards can differ depending on their dialectal background. The considerable complexity of the map of dialects of hilly Slovenia can potentially decrease within-group intelligibility or impede the unidirectional character of spoken comprehension. Additionally, mutual understanding can also differ across generations. Some speakers possess an advantage of being fluent in several contemporary standards, enhanced by the use of Serbo-Croatian amalgam widely spoken in former Yugoslavia (Požgaj Hadži et al. 2013). The Timok-Prizren dialect is treated as a transitional character between Serbo-Croatian group and South-East Slavic vernaculars. It exhibits selected features typical for both groups.

2.4.2 Eastern group

Eastern group of the South Slavic languages is most often described in terms of the erosion of synthetic declination, presence of the post-positive definite article and the lack of infinitive verb form replaced by a subordinate clause *da* + verb, as well as exceptionally rich conjugation paradigms including imperfectum, plusquamperfectum, futurum exactum, futurum praeteriti which are either treated as archaisms or are not in use in the other Slavic languages. Furthermore, the *po-* prefix, used in Macedonian and Bulgarian as comparative marker, can cause a comprehension difficulty, due to the different function of such in other Slavic languages, as well as different paradigms of comparison of adjectives and adverbs. On the phonological level, the Bulgarian and Macedonian group, in the course of its development, preserved several characteristics which distinguish it from the other groupings, such as the lack of epenthetic *l*. Relics of the Proto-Slavic **tj* and **dj*, namely, *št*, *žd* sequences are a common factor for the group. The reflexes of *ě* also create a distinction between South-East from South-West subgroups.

Mutual intelligibility within the group is intact. Linguistic differences within the Eastern group are marginal. Furthermore, effective communication with users of the South-West group is possible due to the common language of former Yugoslavia. Structurally, only limited features differentiate Bulgarian and Macedonian, one of which concerns the word stress distribution patterns. Bulgarian exhibits free stress placement, as opposed to Macedonian which has a fixed antepenultimate word stress. The process of strong *jer* vocalization in the Macedonian area resulted in *o* and *e*, whereas in Bulgarian the shift resulted in (*ъ > ъ, ь > *e*) with several exceptions in the Eastern dialects and Standard Bulgarian. This isogloss is one of the most archaic demarcation lines in South-East

Slavic area. Such a division also finds a confirmation in preserved writings created in two Old Church Slavonic recension areas, that is Preslav and Ohrid. In Bulgarian, nasal **ǫ* altered into ъ (and in peripheral dialects into *a*), whereas in Macedonian mainly in *a* (e.g., Proto-Slavic **sǫsědъ* > Bulgarian *съсед*). In Macedonian *ę* became denasalized and resulted in *e* but also gave *a* in direct vicinity of *j* (e.g., Proto-Slavic **pętь* > Macedonian *pet*, Proto-Slavic **ęzykъ* > Macedonian *jazik*). The process of the reduction of unstressed vowels left its traces in Bulgarian speech. The reflex of *jat* defines the isogloss dividing the Bulgarian area into Western and Eastern vernaculars.

The articulation of *ě* in the Eastern dialectal area of Bulgaria is open. Additionally, in the Southern linguistic territory, open realization occurs regardless of stress, whereas in the North, the open articulation of remnants of *jat* is noted only in stressed positions and when the following syllable lacks the front vowel. This is in contrast to the Western dialectal area, in which *jat* is consequently realized as mid-front segment. The Eastern dialectal territory, composed of North-East and South-East area, is also characterized by the reduction of unstressed vowels as well as consequent palatalization of consonants followed by front vowels. The Macedonian dialects are, in contrast, divided by the vertical axis into Northern, Central, and Southern dialects mainly based on the alternations in development of the Proto-Slavic **tj* and **dj* sequences. Speech of the Ohrid area, which belongs to the Western dialectal territory of Macedonia does not note palatalization process before mid-front vowels in onset (Sławski 1962). The peripheral South dialectal areas of Macedonia, however, exhibit partially free lexical stress as well as reduction of unstressed vowels. Also, in several relatively small areas of Southern vernaculars the nasal sounds are noted. The Northern dialectal territory presents a transitional character and incorporates several features of Serbo-Croatian standards. The Torlakian dialect, with many features common for the *Balkansprachbund* is often classified as transitional.

Old Church Slavonic belongs to the South-East Slavic grouping. It was codified by Saints Cyril and Methodius in ninth century AD based on the South-East dialectal stratum. Its archaic character confirms the law of open syllable. Old Church Slavonic texts still provide an excellent material for diachronic studies on Slavic languages and preserve traces of the late Common Slavic era. The syllable nuclei in Old Church Slavonic could be placed on long vowels (*i*, *y*, *u*, *ě*, *a*); short vowels (ь, ъ, *e*, *o*); nasals (*ę* and *ǫ*) and syllabic consonants (*r*, *l*). Numerous cases of epenthetic *l* have been noted in early Old Church Slavonic (Bartula

2011). Furthermore, the presence of the *št* and *žd* consonantal sequences, as the ancestors of *tj*, *kt'*, and *dj* clusters is common for the Old Church Slavonic. The differentiation of the Old Church Slavonic into separate recensions dates to the twelfth century AD.

2.5 Slavic microlanguages

Alternative classifications, such as the one proposed by (Dulichenko 1981), highlight that apart from either well established or forcefully introduced standards, several minor languages, which typologically belong to the Slavic branch, can be distinguished.

A large diaspora of Slavic language users maintains Slavic languages or koiné, which do not happen to be officially proclaimed state languages. These include: Rusyn or Ruthenian (admitted as a minority language - Carpatho-Rusyn and Pannonian Rusyn of Vojvodina), Burgenland Croatian (minority language spoken in Austria, based on Čakavian stratum), Molise Croatian (spoken in southern Italy with Croatian stratum), or Pomak (spoken by Pomak community in Greece and Turkey based on Bulgarian stratum).

The other classifications suggest that the transitional character of some vernaculars permits to acknowledge them as separate microlanguages, such as the speech of Banat (spoken in Banat, in Serbia and Romania, based on Bulgarian stratum), Resian (Slovene based dialects spoken in Italy), Prekmurian (spoken in Hungary with Slovenian stratum) Bunjevac speech (common in the Bačka region), Lachian (treated as transitional lect between Polish and Czech) or Podlachian (ethnolect spoken in Podlachia region) (Vuković 2015; Kowalski 2015; Samardžija 2015; Stankiewicz 1986; Walczak-Mikołajczakowa 2015).

Often, an extra-linguistic factor, such as a question of identity formation, accompanies the process of vernacular acclamation. It is raised in the discussions on the status of Silesian, Moravian and to a lesser extent Kashubian. Leaving aside the issue of the official recognition, several Slavic languages have been given a status of endangered either by Ethnologue or UNESCO. However, since many speakers of the abovementioned languages, ethnolects, literary dialects or pluricentric vernaculars are often proficient users of another standardized variety, or either bilingual or fluent users of a non-Slavic language, the question of intelligibility between their L1s and phylogenetically more distant languages requires separate investigation.

Since the theoretical foundation of spoken intercomprehension is based on the above-mentioned similarities of the Slavic languages, the following chapters will attempt to examine the role of structural resemblances between the Slavic languages and attempt to examine their impact on mutual intelligibility. Therefore, such a comparative description of features shared by the tested languages will be further validated on the basis of the empirical studies into receptive multilingualism.

Chapter 3

Identification of Linguistic Origins

The structuralist narrative which binds the conducted experiments suggests starting from the smallest distinguishable units of language. Since phonemes canonically do not convey meaning, with the exception of onomatopoeic phrases, the study which opens the empirical part of this work is devoted to segmental and suprasegmental analyses of logatomes and their importance in identifying linguistic origins of speakers. This chapter raises the question of lay listeners' competence in quasi-forensic investigation. The chapter addresses the questions of three dimensional vowel overlap as well as stress distribution patterns and their role in the identification of speakers' native language. It stresses the importance of acoustic cues in spoken language analyses for the determination of the origin.

The manuscript summarizing the study is currently in the review process. The experimental results were first presented to the public during the Speech Science Workshop organized jointly by Zurich University and Saarland University on 29. January 2021.

3.1 Introduction

Spoken language identification (henceforth: LID) is a complex process of perceptual recognition or automatic identification of a language from a spoken sample. Recently, fine-grained LID, including areal and dialectological investigations have become widely applicable in a procedure known as LADO: Language Analysis for the Determination of Origin (Patrick 2012). Also in forensics, automatic and human-based LID is an important stage of spoken evidence validation and speaker profiling. In the present work, a human-based auditory language identification task was conducted using spoken material from four Slavic languages: Bulgarian, Czech, Polish, and Russian. The study combines several methodological components. Firstly, the acoustic-phonetic component provides an explanation of cross-linguistic influence of vowel space overlap on test performance. Secondly, a Slavic component explains the role of listeners' L1 in the recognition of linguistic origins of Slavic speakers. Thirdly, an information theory approach attempts to quantify cues in LID; and further, the perceptual LID results were compared with the information-theoretic notion of surprisal.

3.1.1 Aims and hypotheses

The main goals of this work are: (1) to define a minimal segmental and suprasegmental piece of information required for correct recognition of speakers' L1 in human-based LID; (2) to correlate the phonetic and typological similarities of participants' L1 and the respective stimulus with their test performance; (3) to investigate the relation of the three-dimensional vowel overlap measured by means of Pillai score (Hay et al. 2006; Nycz and Hall-Lew 2013) with the subjects' performance in LID tests; and (4) to examine the effect of sound variation on lay listeners' ability of language identification from an information-theoretic perspective (Shannon 1948).

It is assumed that (1) lay listeners whose L1 is closely related to the language of the recording can correctly identify the linguistic origin of speakers in the auditory modality even when exposed to logatomes; (2) the alternations in stress position and length of the logatomes influence the recognition scores, with better performance on longer logatomes; (3) spectral characteristics of the signal, such as cross-linguistic vowel space and duration overlap, are correlated with human performance in LID tests; and (4) the mean logatome identification surprisal (LIS) values between the tested languages are correlated with the experimental results, that is, lower LIS values impede linguistic discrimination. Hence, it is expected that the lower the mean LIS values between two languages, the higher the LID scores. Moreover, it is worth investigating which phonetic features and their cross-lingual overlap improve listeners' performance. What is the optimal threshold of phonetic and acoustic information required for correct identification of speakers' origins? Do suprasegmental features and language-specific word stress distribution play a role in LID tasks? How can the information-theoretic notion of surprisal help us determine and predict human performance? These issues will be addressed in this study with a focus on spectral and temporal properties of segments that are common in the phonological inventories of the Slavic languages.

3.1.2 Related work

The scientific debate on the relevance of non-linguist native speakers in LADO or LOID (Linguistic Origin Identification) is vibrant, and polarized with arguments both in favor of such an approach when supervised by professionals (Cambier-Langeveld 2016; Nolan 2012; Wilson and Foulkes 2014), and against the involvement of lay listeners altogether (Eades 2005; H. Fraser 2011; Patrick 2010). The arguments for the involvement of native speakers are the lack of independent supervision of experts conducting LADO, as well as monopolist practices. Furthermore, the question of reliable counter-expertise contributes to the engagement of lay listeners in the task of L1 identification. The lack of academic literature on a particular vernacular can justify the engagement of non-trained native speakers in the identification of linguistic origin. Furthermore, a comparison of the accuracy of LADO professionals, academic phoneticians, phonetics students, and untrained native speakers has indicated that the last group performed best, whereas the LADO professionals' results were at chance level only (Foulkes and Wilson 2011). In contrast, the arguments for the exclusion of native speakers in LADO procedures touch upon non-experts' unfamiliarity with the proper terminology, or their possible bias in providing an opinion. Additionally, there is high variability in listener accuracy and it is difficult to predict which listeners might perform better than others. Fraser has pointed out (H. Fraser 2009) that untrained listeners often fail to identify dialects correctly. Since trained professionals and native speakers often pay attention to different cues and may have developed different skills due to training versus natural acquisition, a collaborative involvement of both parties seems to be an ideal solution.

It is also interesting to explore the relation of linguistic fluency and performance of lay listeners in LID tests, as well as the phonetic and phonological correspondences between listeners' L1 and perceived speech. The importance of phonological insights in auditory LID has previously been stressed by (Peperkamp and Dupoux 2002) who suggested that listeners' sensitivity to stress cues depends on the stress function in their native language. Even when a particular feature associated with word stress is not phonemic in participants' L1 and does not play a role in lexical discrimination in one's own language, listeners might still be more sensitive to stress distinctions in a foreign language. The effect of stress predictability was also reported in the perception of non-words for speakers of languages exhibiting phonemic stress. Regarding the participants in this study, a weak 'stress deafness' has previously been observed among Polish native speakers (Peperkamp et al. 2010).

Regardless of the uneven functional load of the units used in LID tasks, several studies have confirmed the possibility of identifying a language when exposed to limited information and distorted speech. Hence, apart from the inherent phonetic features, syllable structure can also serve as a cue in LID, especially in Slavic languages, which are known to exhibit complex onset clusters such as CCCCVC, in both Polish and Russian. Information conveyed in phonotactic rules of one language can intuitively lead to successful language identification by humans as well as by automatic LID systems (Navratil 2001; Zissman and Berkling 2001).

Prosodic cues are similarly meaningful in auditory language identification. Several techniques of limiting the spectral information in perceptual LID tasks have previously been applied, including spectral envelope removal and temporal-envelope modulation. Studies involving modifications of speech signals have confirmed that prosodic cues play an important role in LID, even when separated from the segmental information (Ohala and Gilbert 1979). Even though prosody is rarely, or never, decisive in contemporary LOID assessments (Hoskin 2018), research on suprasegmental cues in LID has led to the proposal of a rhythmic model of language identification by (Rouas et al. 2005). Regardless of signal distortion, the accuracy of language recognition still varies depending on the source data type: read or spoken, with higher accuracy on the former. For instance, a domain-dependency in a machine LID on the same set of Slavic languages was recently reported by (Abdullah et al. 2020).

3.1.3 LADO and auditory language identification

Perceptual language recognition is a complex operation, involving several cognitive processes. It is a multidimensional action without discrete component stages, in which graded information flows in a cascade (McQueen et al. 2011). This process is based on various sources of acoustic, linguistic, and extra-linguistic information. Perceptual as well as machine-based LID techniques can be improved with training. Studies in contemporary cognitive linguistics have shown that the ability of humans to identify a language can be significantly improved by training or exposure to a particular language or vernacular, similarly to automatic LID, in which the size of training data often predicts overall system performance. Both methods, perceptual and machine-based LID, are currently widely applied not only for commercial purposes, such as translation and localization, but also in the field of jurisdiction and forensics. Governments and intelligence bureaus often take advantage of associated or external agencies to perform LID for the needs of criminal investigations. Furthermore, LADO tests are widely used to verify the origin of asylum seekers, by closely investigating their speech using the methodological apparatus of acoustic phonetics, linguistics, and dialectology. In such a procedure, there is an underlying assumption of an existing connection between how people speak and their ethnic or national origin, or rather the place of their socialization, which is a natural consequence of language acquisition in a certain linguistic community. Nevertheless, lay listeners' expertise is often valuable in linguistic, or more precisely, dialectal background evaluation. 'Guidelines for the use of language analysis in relation to national origin in refugee cases' suggest that native speakers should not be treated as experts when evaluating a speaker's origin on the basis of their speech (Language and National Origin Group (LANG 2004)). On the other hand, especially for cases involving languages with limited digital resources, agencies specialized in conducting LADO or LOID tests often make use of native speakers in cooperation with linguists to estimate the target's place of origin (Hoskin 2018). Such a practice is partly pragmatic, having in some cases to do with lack of descriptions or linguists specializing in the languages concerned. This is also founded on the widely accepted principle that native speakers are in general the most knowledgeable informants.

The so-called 'intelligent guessing' (Meissner 2018) of a language depends on listeners' linguistic and extra-linguistic knowledge, exposure to other languages, associative competence, as well as short-term and long-term memory storage and, above all, the quality and quantity of information perceived in the process of recognition. It has been shown that, even when a signal for auditory identification is highly degraded, subjects are aware of the cues they perceive in LID tasks (Muthusamy et al. 1994). Furthermore, information provided in LID itself can be limited to a particular subsystem of a language (Schultz and Waibel 1998). For instance, it is possible to distinguish two languages based solely on the presence of a phone in one language and its absence in another (Harper and M. Maxwell 2008). Strictly phonetic knowledge in LID based on the characteristics of vowel systems was proposed by Pellegrino and André-Obrecht (Pellegrino and André-Obrecht 2000). Articulatory classes were also investigated as delimitation features in language recognition (Kirchhoff and Parandekar 2001). In this application, several distinguishing labels were defined, for example, manner of articulation, consonantal place of articulation, vocalic place of articulation, lip rounding, front-back tongue position, voicing, and nasality. Phonotactic rules in combination with labeling of broad phonetic classes can also constitute a kernel of the language identification process. This approach was proposed by House and Neuburg in the 1970s (House and Neuburg 1977). More fine-grained analysis involving strictly acoustic signal characteristics seem to be of relevance to automatic LID. Perceptual behavioral studies concerning acoustic data and LID are canonically gravitating towards a correlation of characteristics of formant dynamics, voice onset time (VOT) and center of gravity (CoG) along with their fluctuations in the signal with language-specific data. Hence, it seems interesting whether there is a cross-linguistic correlation of spectral and temporal features of vowel systems with the performance in auditory recognition of speakers' origins.

3.2 Methods

The task was presented as a game in which subjects played the role of investigators in a bank robbery case. Their task was to identify the origin of a speaker by listening to an artificial language made up by the speakers-suspects to mislead the investigators. This setup provides the rationale for the application of pseudowords in the LID sessions and drew the participants' attention to non-lexical cues.

3.2.1 Speakers

In total, 40 native speakers (5 males and 5 females per each tested language) of Bulgarian, Czech, Polish, and Russian were asked to read a list of nonsense disyllabic CVCV and trisyllabic CVCVCV items. Based on a questionnaire distributed after the recording sessions, the speakers whose voices were selected for use in the experiment were profiled as users of the standard language variety, with middle socioeconomic status, having completed secondary education or currently enrolled at a university program, aged 21-36, and having experienced no surgical operations in the ear, nose, and throat region nor having required speech or hearing therapy.

3.2.2 Participants

In total, 228 speakers of four Slavic languages (50 Bulgarian, 53 Czech, 66 Polish, and 59 Russian speakers) participated in the task. Before the experiment, the subjects filled out a questionnaire with basic demographic information and questions about their linguistic background. The participants were asked about their linguistic proficiency (CEFR scale), multilingualism within their family, their language of everyday communication, education background, and years spent abroad. None of the participants reported any hearing difficulties. Since previous studies revealed that multilingualism significantly influences performance in identifying an unfamiliar language (Muthusamy et al. 1994), the results were post hoc correlated with the linguistic profiles from the questionnaires. The participants had no formal training in other Slavic language. Data obtained from subjects having background in Slavistics, linguistics, forensics, or phonetics ($< 1\%$) were excluded from the analysis.

3.2.3 Design of materials

To avoid the possibility of overlapping with existing lexemes of Slavic languages and associating with meaningful tokens, the NUP (nonword uniqueness point) threshold had to be achieved for every item presented in the identification task (Cutler 2012). Furthermore, the stimuli contained only vowels and consonants that are present in the phonological inventories of all four investigated languages. Stimuli with a fixed stress position were created in line with the natural stress distribution rules in these languages. In order to avoid the effect of unnatural articulation and audibly perceptible reading difficulty, the set of stimuli consisted only of only sequences that yielded effortless and natural pronunciation. The pseudowords were marked as easy to read by the speakers, which ensured effortless articulation and natural rendering of the spoken samples (Bonatti et al. 2005).

The set of stimuli was created according to the following rules: (1) all logatomes were in line with the open-syllable principle, the common law in Slavic before the vocalizations of the semivowels; (2) the items consisted of stops /k/, /g/, /p/, /b/, /t/, /d/ and a combination of five common vowels /a/ /ɛ/, /i/, /ɔ/, /u/, which was justified by the degree of interference of plosives with the adjacent vocalic segment (Stevens and House 1963; Ladefoged and Maddieson 1996) as well as by the results from previous studies using logatomes in perceptual discrimination, which showed that the diphones 'pa', 'si' and 'ki' are among the most discriminable elements; (3) to control for a fixed stress position and length of the pseudowords, both bisyllabic CVCV and trisyllabic CVCVCV sequences were used in the test; (4) no zero-onset was present in the tokens, even though this structure is possible in all investigated languages; (5) primarily non-palatalized segments were used due to the unequal distribution and frequency of palatalized CV sequences in the investigated languages; (6) in order to avoid a priming effect, only non-nasalized units were taken into consideration due to frequent synchronous and asynchronous nasals in Polish as opposed to other Slavic languages.

3.2.4 Recording procedure

A sex-balanced group of 40 speakers of the Slavic languages was given a self-paced reading task in an acoustically controlled setting. The readers were native speakers of one of the four Slavic languages, who evaluated themselves as users of the standard variety of the language. To avoid possible uncontrolled effects and paralinguistic distractors, such as recording-related anxiety causing different speech rates, coughing, etc., each list of pseudowords was read and recorded twice, which resulted in 4000 tokens (40 readers x 50 tokens x 2 sessions). The recordings were randomized and intervals between the tokens were standardized. These samples were used as the audio stimuli in the LID task.

3.2.5 Speech analysis procedure

Before the experiment, the stimuli were automatically segmented and annotated using the BAS (Kisler et al. 2017), visually inspected and, if necessary, manually corrected in Praat (Boersma and Weenink 2020). The F1 and F2 values in critical band rate (Bark) were extracted from the midpoint of the vowels using LPC Burg's method. The results of these estimations were visually inspected and compared with the corresponding wideband FFT spectra. The manual verification and alignment with the nearest zero-crossing correction allowed for precise duration measures [in ms]. Then, the vowel space overlap of the 5 investigated vowels was calculated by means of the Pillai method (Hay et al. 2006; Nycz and Hall-Lew 2013) as a multivariate analog of the F-ratio from the analysis of variance based on ANOVA (Pillai 1954; Bray and S. E. Maxwell 1985). As suggested by recent results from the Monte Carlo simulation, this method outperforms the other vowel overlap measures such as SOAM (spectral overlap assessment metric), VOACH (vowel overlap analysis with convex hulls) or Euclidean distance between centroids of respective vowel spaces (Kelley and Tucker 2020). Furthermore, computed Pillai scores also take into consideration a three-dimensional plane including the duration of segments, a factor highly relevant in the LID process and important when languages with significant differences in vowel lengths are considered.

3.2.6 Experimental procedure

The first stage of the experiment examined the acoustic properties of segments in the spectral and temporal domains. The second phase introduced the fixed versus free stress-position factor, by implementing more complex stimuli with variations in stress positioning: namely, on the initial and penultimate syllables versus free stress distribution. The randomized disyllabic and trisyllabic logatomes were binaurally presented to the subjects. In order to exclude any orthographic influence, no visual input was given. The participants were provided with a controlled amount of acoustic information, excluding semantics. They were allowed to listen to the samples three times. Before the identification task began, subjects were given the opportunity to practice on trial samples, at which time they could adjust the volume to a comfortable level. During the entire session, participants were exposed to 64 randomized samples containing long and short logatomes. After each sample, they were asked to identify the language of the speaker by choosing from a closed set of

Slavic languages: Bulgarian, Czech, Polish, and Russian. This setup corresponds to the speaker verification task, a common practice in LADO procedures. A 5-point Likert scale was used to rate the subjects' confidence in their choice. After having completed the entire session, the subjects were presented with their overall accuracy scores.

3.2.7 Analysis

A language confusion matrix was created from the results of the auditory LID task. Afterwards, a similarity index (SI) of the investigated Slavic languages was calculated (Johnson 2003). The perceptual similarity index (PSI) of Bulgarian, Czech, Polish, and Russian was calculated from the similarity scores (Thomas 2010). In the next step, the performance of lay listeners was correlated with acoustic segmental similarity measures by means of the 3D Pillai scores across the four languages. The results of the Slavic language of origin recognition task were compared afterwards with the amount of acoustic and phonetic information given in the samples. For the purposes of the Logatome Identification Surprisal (LIS) calculation, the logatomes were narrowly transcribed in the IPA 2020 standard and were then paired with their equivalents in the other three languages. As a result, 19164 logatome pairs were transcribed (3283 Bulgarian-Czech logatome pairs, 3305 Bulgarian-Polish, 3304 Bulgarian-Russian, 3084 Czech-Polish, 3083 Czech-Russian, and 3105 Polish-Russian). The numbers are different for each language pair due to some mispronunciations by the readers, yielding new logatomes with no equivalents. A logistic regression analysis was conducted to address the research question regarding differences in recognition based on logatome length. Then, a Pearson's χ^2 test was conducted to test the relation between Slavic L1 and participants' performance, and to correlate the logatome identification surprisal with the similarity scores. The performance scores were correlated with the subjects' linguistic profiles in terms of their mother tongue and fluency in non-Slavic languages. Language similarity sequences for disyllabic and trisyllabic logatomes were compared with respect to stress distribution patterns in Bulgarian, Czech, Polish, and Russian. Finally, the results in the language confusion matrix (the LID scores) were correlated with the mean logatome identification surprisal (LIS) values for each language pair, in order to investigate the effect of regularities in cross-linguistic sound correspondences on listeners' performance in auditory identification of language of origin.

3.3 Results

The following sections introduce the results of the experiment according to each methodological component.

3.3.1 Perceptual similarity score

The confusion matrix generated from the results of the language identification task served as the basis for the calculation of the perceptual similarity indices (PSI). The similarity scores obtained from the overall performance were as follows: for Polish and Russian, 0.25; for Czech and Polish, 0.48; for Bulgarian and Czech, 0.52; for Polish and Bulgarian, 0.55; for Czech and Russian, 0.21; and for Bulgarian and Russian, 0.47. The pairwise perceptual similarity index from the overall performance was calculated as the following equation:

$$PSI = -ln(Pxy + Pyx/Pxx + Pyy)$$

in which Pxy stands for proportion of language X identified as language Y, Pyx equals to proportion of language Y identified as language X, Pxx - proportion of language X identified as X, and Pyy - proportion of language Y identified as Y. The results are illustrated in the network plot (Figure 3.1). The distance between language nodes in the network plot reflects the perceptual similarity scores (PSI) between the languages. The network plot was created by projecting language nodes into a 2-dimensional space using a force-directed graph function and multidimensional scaling, based upon the perceptual similarity distance scores yielded from the experiment. The colors of the language nodes themselves represent the genealogical groupings of the languages within the Slavic family (green – South Slavic, blue – West Slavic, red – East Slavic). The alternations in similarity scores between the disyllabic and trisyllabic sequences are presented in Figure 3.3.

3.3.2 Logatome length and stress position

A logistic regression analysis was conducted to address the research question regarding the differences in identification based on logatome length. A threshold of statistical significance $\alpha = 0.05$ was applied. The hypothesis of more accurate recognition of longer trisyllabic sequences with respect to disyllabic logatomes was rejected ($\chi^2 = 2.28; p = 0.131$). The

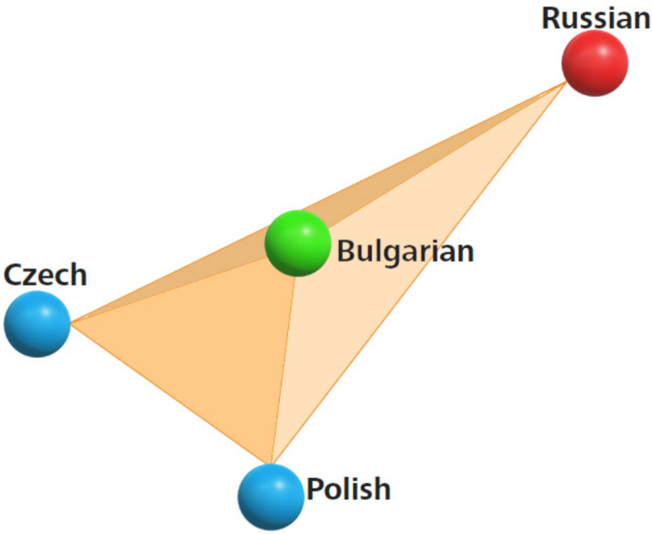

Figure 3.1: Distance between languages based on perceptual similarity scores. Distance between the nodes corresponds to the PSI yielded from the experiment. The edges connecting the language nodes are scaled according to the degree of similarity of each pair of languages.

length of the logatomes was not a relevant predictor of correct identification, Cox and Snell's $R^2 < 0.01$. The correct identifications of disyllabic stimuli reached 46.8% (n = 3397); whereas the accuracy of recognition after CVCVCV sequences was 45.6% (n = 3304). Detailed results are presented in Table 3.1. Nuances in stimuli length resulted in different clustering based on subjects' performance. The divergence of Russian compared with the other three Slavic languages is constant among the sequences. The similarity of Bulgarian, Czech, and Polish, however, differs with respect to the length of the logatomes. The results from the overall performance indicate that Polish is perceptually closer to Bulgarian than to Czech. A closer examination of the scores with respect to the length of the pseudowords reveals alternations between Polish and Bulgarian, and shows a greater similarity between Czech and Bulgarian instead (see Figure 3.2).

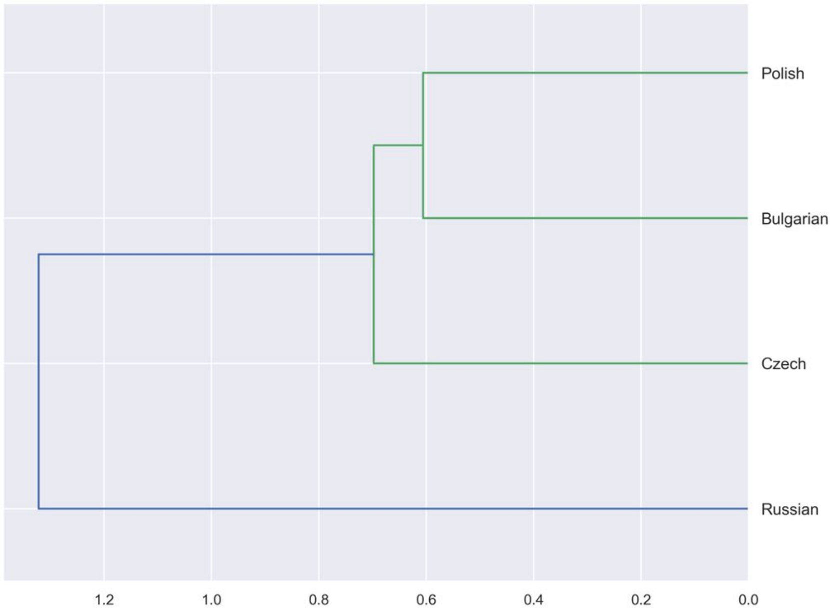

Figure 3.2: Dendrogram clustering language similarities based on subjects' performance. Language clustering based on the overall results.

3.3.3 Native Slavic language and non-Slavic L2

Test participants were native speakers of four Slavic languages: Bulgarian, Czech, Polish, and Russian. Their fluency in non-Slavic languages was documented in the questionnaire. Interestingly, fluency in non-Slavic languages also appeared to be a predictor of their LID performance. The results showed that the model fits the research question well ($\chi^2 = 126.18; p < 0.001$; Cox and Snell's $R^2 = 0.009$).

Apart from Slavic languages, fluency in Romanian and Azeri appeared to influence the LID scores, even though their systems of monophthongs differ from those of Slavic languages. L1 speakers of Polish had 39% higher correct identification, Odds Ratio = 1.39. Fluency in Czech increased performance by 41%, $OR = 1.41$, whereas knowledge of Russian improved the accuracy by 17%, $OR = 1.17$. The effect of fluency in Bulgarian did not reach the threshold of statistical significance. Regarding languages that are typologically unrelated to Slavic languages, fluency in Romanian increased the performance in the LID task 1.6 times,

Table 3.1: Results of logistic regression analysis for correct responses depending on logatome length

	B	SE	$Z(1)$	p	OR	95% CI for OR	
						LL	UL
Stim. length	0.05	0.03	2.28	0.131	1.05	0.99	1.12
Const.	−0.18	0.02	56.37	< 0.001	0.84		

B – unstandardized coefficient, SE – standard error, Z – Wald test, p – p-value, OR – odds ratio, CI – confidence interval, LL – lower limit, UL – upper limit

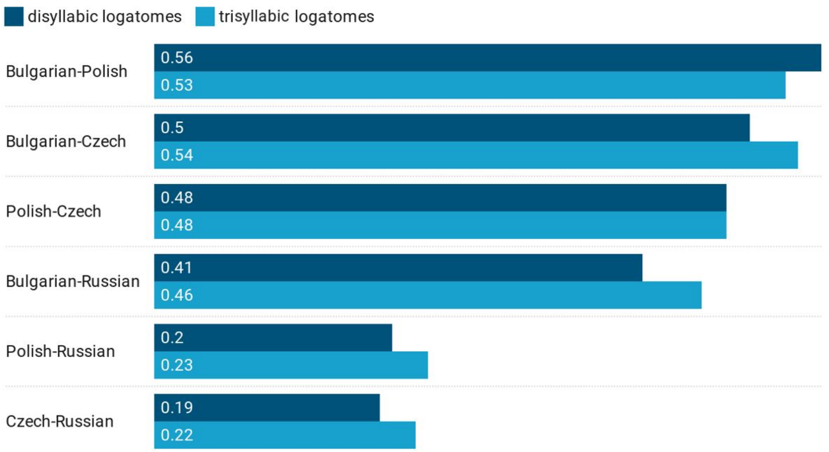

Figure 3.3: Pairwise similarity scores for disyllabic and trisyllabic logatomes

$OR = 2.60$, whereas the fluency in Azeri lowered the performance by $77\%, OR = 0.23$. Results including all languages are presented in Table 3.2.

3.3.4 Native Slavic language and identification scores

In the next step, the relation between listeners' Slavic L1 and accuracy of identification was investigated by means of a Pearson's χ^2 test. The

Table 3.2: Results of logistic regression analysis for all subjects' languages

						95% *CI* for *OR*	
	B	SE	$Z(1)$	p	OR	LL	UL
Hebrew	0.18	0.17	1.23	0.268	1.20	0.87	1.66
Swedish	0.68	0.44	2.38	0.123	1.98	0.83	4.73
Belarusian	0.59	0.54	1.18	0.276	1.80	0.62	5.21
Norwegian	0.60	0.31	3.79	0.052	1.82	1.00	3.33
Spanish	0.01	0.16	0.01	0.928	1.01	0.74	1.39
Portuguese	0.18	0.19	0.84	0.358	1.19	0.82	1.74
Greek	0.10	0.27	0.14	0.713	1.10	0.66	1.85
Serbian	0.36	0.37	0.94	0.331	1.43	0.70	2.92
Croatian	−0.19	0.44	0.20	0.656	0.82	0.35	1.93
Polish	0.33	0.08	16.26	< 0.001	1.39	1.19	1.64
Japanese	0.24	0.27	0.78	0.377	1.27	0.75	2.15
Romanian	0.95	0.24	0.78	0.377	1.27	0.75	2.15
Ukrainian	−0.30	0.28	1.20	0.273	0.74	0.43	1.27
English	0.04	0.11	0.14	0.711	1.04	0.85	1.28
Italian	−0.09	0.17	0.25	0.615	0.92	0.66	1.29
Finnish	0.33	0.27	1.52	0.218	1.40	0.82	2.37
French	−0.17	0.19	0.84	0.360	0.84	0.59	1.21
German	−0.01	0.10	0.01	0.940	0.99	0.81	1.21
Czech	0.34	0.08	18.15	0.001	1.41	1.20	1.65
Russian	0.16	0.08	3.99	0.046	1.17	1.00	1.37
Afrikaans	0.10	0.20	0.25	0.617	1.10	0.75	1.62
Bulgarian	0.10	0.09	1.48	0.224	1.11	0.94	1.31
Const.	−0.41	0.08	28.48	< 0.001	0.66		

B – unstandardized coefficient, SE – standard error, Z – Wald test, p – p-value, OR – odds ratio, CI – confidence interval, LL – lower limit, UL – upper limit

Table 3.3: Overall scores with regard to Slavic L1

Native language		Response language							
		Bulgarian		Czech		Polish		Russian	
		n	%	n	%	n	%	n	%
Bulgarian	-	$395_{a,b}$	53.7	374_b	50.8	487_c	66.2	$441_{a,c}$	59.9
	+	$341_{a,b}$	46.3	362_b	49.2	249_c	33.9	$295_{a,c}$	40.1
Czech	-	580_a	65.9	296_b	33.7	540_a	61.4	418_c	47.5
	+	300_a	34.1	583_b	66.3	340_a	38.6	462_c	52.5
Polish	-	704_a	64.1	629_b	56.7	431_c	39.3	473_c	43.2
	+	394_a	35.9	48.1_b	43.3	667_c	60.7	621_c	56.8
Russian	-	559_a	65.7	502_b	58.6	577_a	67.6	254_c	29.8
	+	292_a	34.3	355_b	41.4	277_a	32.4	599_c	70.2

Indexes a, b, c – differences on the level $p < 0.05$ with Bonferroni correction, +/- correct/incorrect identification

analysis revealed statistically significant results for each group of respondents. The scores are presented in Table 3.3. In the group of Bulgarian native speakers, the best recognition scores were observed for Czech ($\chi^2(3) = 42.29; p < 0.001; V = 0.12$). Polish was identified in 33.8% of cases; correct identifications of Russian reached 40.1%; whereas the group L1 identification scores, 46.3%, were similar to Czech and Russian. In the group of Czech native speakers, the highest discrimination scores, 66.3%, were measured for Czech ($\chi^2(3) = 224.55; p < 0.001; V = 0.25$). Less accurate identifications were observed for speakers whose L1 was Russian 52.5%; Polish, 38.6%; and Bulgarian, 34.2%. Polish native speakers accurately recognized speakers' dominant language as Polish in 60.7% of cases ($\chi^2(3) = 176.81; p < 0,001; V = 0,20$). The second most correctly identified language was Russian, 56.8%. The recognition scores of Bulgarian and Czech ranged from 36% to 42%. The native speakers of Russian identified Russian in 70.2% of ($\chi^2(3) = 317.75; p < 0,001; V = 0.31$) cases. The least accurately recognized linguistic origins were Czech, 41.4%; Bulgarian, 34.3%; and Polish, 32.4%.

3.3.5 Acoustic and perceptual measures

Pillai scores were computed to discover the relation between vowel overlap and performance in language identification. The results in Table 3.4 correspond to the cross-lingual overlap of five vowels calculated using

Table 3.4: Vowel overlap in Pillai scores for all tested vowels

Language pair	Vowels				
	a	e	i	o	u
Czech-Polish	0.031	0.075	0.100	0.032	0.031
Czech-Russian	0.017	0.060	0.041	0.013	0.100
Czech-Bulgarian	0.015	0.359	0.032	0.058	0.043
Polish-Russian	0.071	0.202	0.069	0.026	0.055
Polish-Bulgarian	0.063	0.488	0.006	0.059	0.037
Russian-Bulgarian	0.045	0.163	0.025	0.017	0.027

the Pillai method, including durations of segments and their F1 and F2 values. The Pillai scores reflect the three-dimensional similarity of vocalic segments in Bulgarian, Czech, Polish, and Russian. Higher Pillai scores indicate more diverged (less overlapping) segments. The results show that the overlap of low-central /a/ tokens was not correlated with correct language recognition; whereas types of /ɛ/ ($OR = 0.43$) and /u/ ($OR = 52.97$) were highly correlated with Pillai scores. More divergent vowels correlated with greater accuracy in the recognition of the language it belongs to. The detailed results are given in Table 3.5. Although vowel space is rarely mentioned as discriminable on the word level in classical models of speech processing, it appears to have had an influence when sample discrimination is performed on the basis of highly distorted signal.

3.3.6 Perceptual measures: Logatome identification

As suggested by (Skirgård et al. 2017), languages more different in sound appear easier to distinguish than the more similar ones. An information theoretic notion of surprisal is one method for quantifying these differences. In this context, surprisal metrics quantify the informativity of cross-linguistic unit correspondences in bits. In the current LADO setting, sound identification surprisal (SIS) is computed according to the following equation:

$$SIS(L1 = s1|L2 = s2) = -log_2 P(L1 = s1|L2 = s2)$$

Table 3.5: Vowel overlap in Pillai score and subjects' performance

	B	SE	$Z(1)$	p	OR	95% CI for OR LL	UL	R^2	χ^2
a	−0.50	0.46	1.19	0.276	0.61	0.25	1.49	< 0.01	1.11
const.	−0.35	0.05	48.34	< 0.001	0.71				
e	−0.85	0.18	21.13	< 0.001	0.43	0.30	0.62	< 0.01	21.26***
const.	−0.16	0.05	8.42	0.004	0.86				
i	1.00	0.56	3.31	0.077	2.72	0.90	8.21	< 0.01	3.14
const.	−0.51	0.06	83.87	< 0.001	0.60				
o	−0.86	0.65	1.74	0.187	0.42	0.12	1.52	< 0.01	1.74
const.	−0.25	0.06	15.95	< 0.001	0.78				
u	3.97	0.68	34.58	< 0.001	52.97	14.11	198.89	0.01	34.75***
const.	−0.67	0.07	89.47	< 0.001	0.51				

B – unstandardized coefficient, *SE* – standard error, *Z* – Wald test, *p* – p-value, *OR* – odds ratio, *CI* – confidence interval, *LL* – lower limit, *UL* – upper limit, R^2 - Cox and Snell's index, χ^2 - Pearson's chi-squared test

in which *L1* stands for response (decoder) language, *s1* corresponds to sound of the response (decoder) language, *L2* – exposure (stimulus) language, and *s2* – sound of the exposure (stimulus) language.

The SIS values obtained for combinations of stimulus and decoder languages allowed the quantification of the (un)expectedness of cross-linguistic sound correspondences and of the pseudoword pairs used in the experiment. The overall logatome identification surprisal (LIS) is calculated as the sum of the sound identification surprisal. This provides a quantification of the overall (un)expectedness of each logatome's phonetic form given a corresponding logatome. Since LIS between two logatomes is computed by summing up the SIS values of the sounds from the aligned logatome pair, it depends on the number of available tokens. The normalized LIS values in Table 3.6 were obtained by dividing the LIS value by the length of the logatome. An important property of surprisal-based modeling is that it can reveal asymmetries in the overall identification difficulties depending on the direction of processing. This means that the Bulgarian–Polish exposure–response logatome pair may not have the same normalized LIS value as the Polish–Bulgarian pseudoword pair. To examine possible confusion asymmetries, the mean LIS values between the tested languages were calculated using the incom.py toolbox (Mosbach et al. 2019) and presented in Table 3.6.

Table 3.6: Mean logatome identification surprisal values in bits

Exposure	Response	All sequences	CVCV	CVCVCV
Bulgarian	Czech	0.43342744	0.42931416	0.40465307
Bulgarian	Polish	0.24398879	0.22550414	0.25974784
Bulgarian	Russian	0.62378925	0.42195107	0.43798181
Czech	Bulgarian	0.44885918	0.42195107	0.43798181
Czech	Polish	0.32994016	0.31232876	0.34074236
Czech	Russian	0.32994016	0.31232876	0.34074236
Polish	Bulgarian	0.41876436	0.40292569	0.41646455
Polish	Czech	0.48768054	0.49781193	0.46237080
Polish	Russian	0.61333058	0.55175746	0.64546872
Russian	Bulgarian	0.40262913	0.38082098	0.40771237
Russian	Czech	0.51112285	0.49624590	0.50177142
Russian	Polish	0.21727449	0.16997498	0.25811850

First, the LID scores were correlated with the LIS values between pairs of exposure–response languages for all logatomes, regardless of length. A negative but not significant correlation was found: the lower the surprisal values, the larger the LID scores, (Pearson's $r = -0.36; R^2 = 0.13; p = 0.25$). Furthermore, the mean LIS values for CVCV and CVCVCV logatomes, respectively, were correlated with the subject's performance for CVCV and CVCVCV sequences. The negative correlation for CVCV logatomes (Pearson's $r = -0.40; R^2 = 0.16; p = 0.20$), was found to be stronger than for CVCVCV logatomes (Pearson's $r = -0.27; R^2 = 0.07; p = 0.39$), but neither correlation reached the threshold of statistical significance.

3.4 Discussion

In this study, the ability to recognize a speakers' linguistic origin was investigated. The analysis focused on the spectral and temporal properties of segments shared by the phonological inventories of four Slavic languages: Bulgarian, Czech, Polish, and Russian. In this experimental setup, methodologies from acoustics, phonetics, and information theory were combined to discover which cues (language-specific stress distribution patterns, vowel space overlap, pseudoword length, or logatome identification surprisal) are relevant to lay listeners for determining a speaker's L1. Additionally, an importance of fluency in non-Slavic languages was evaluated. The fixed versus flexible word stress did not appear to be informative enough for the speakers of Slavic languages to influence their performance in the perceptual task. The data suggest that stress distribution is not a discriminable factor. Interestingly, Polish (with a fixed penultimate word stress) and Czech (with a fixed initial word stress) are placed in different groups depending on the stress distribution, which can suggest that stress distribution patters are not informative enough in the L1 identification task. Polish and Czech also did not pattern in the same group according to perceptual similarity, despite their typological proximity – instead, Polish was clustered with Bulgarian. This finding can be explained by significant differences in vowel length among the languages. Czech exhibits vowel lengthening, which distinguishes it from Polish and Bulgarian and hence results in an alternative taxonomy. Furthermore, participants' fluency in a typologically related language significantly influenced their performance in the presented LID task.

3.4.1 Linguistic repertoire effect

Regarding the subjects' linguistic repertoire, the genetic relatedness of their L1 to the stimulus was, intuitively, a valid and strong predictor of performance in the L1 identification task; however, fluency in other non-Slavic languages also had an impact on the LID scores, though, in line with expectations, with a lower effect size. A closer look at the phonetic inventories of the non-Slavic languages correlating with test performance reveals that the interpretation of this finding should be limited to general linguistic knowledge and intuition of lay listeners, rather than viewed as a transfer or mapping between the phonological units of non-Slavic L2 to Slavic L1. The improvement of performance related to fluency in Romanian may be attributed to areal linguistics, as an effect of contact with Slavic languages and its membership in the *Balkansprachbund*. In contrast, fluency in Azeri, a Turkic language, which appeared to diminish recognition scores, may be attributed to the distinctiveness of its vowel inventory, which might result in decreased sensitivity to particular vowel characteristics. The analysis of correlations of Slavic L1 with LID performance demonstrates that listeners were often able to correctly identify the origin of speakers whose L1 was the same as their own, with the exception of the Bulgarian group. The Bulgarian native speakers correctly identified speakers whose L1 was Czech more often than fellow

Bulgarian L1 speakers. This finding can be related to a dialectological landscape of Bulgarian marked by the West-East isogloss, which defines the quality of the reflex of the Proto-Slavic *jat* vowel. Such a difference in vowel quality can cause confusion and thus influence the performance in an identification task in which vowel quality serves as a primary cue.

3.4.2 Vowel space overlap and stress position

Overall, it can be concluded that even without lexical, semantic, or syntactic information, the identification of a speaker's linguistic origin is nevertheless possible based exclusively on the phonetic and phonotactic subsystems. As shown in this study, the 'optimized deduction' or 'intelligent guessing' strategies were confirmed on the material of four Slavic languages. Furthermore, it appeared that it is possible to identify the language of a speaker's origin when exposed to delexicalized audio stimuli as long as the subject's native language and the stimulus language are phonetically close. At the same time, the data did not support the hypothesis of differences caused by disyllabic and trisyllabic sequences exhibiting various stress patterns. Interestingly, an overlap in a three-dimensional vowel space appeared to be a valid predictor of lay listeners' performance. On the other hand, the correlations between the LIS values and LID scores were negative, but low and insignificant. Eventually, other acoustic dimensions such as formant dynamics should be taken into consideration in future studies concerning correlations of signal properties with subject's performance in the LID tasks. The surprisal-based results did not reach the threshold of statistical significance; hence the formulated hypothesis could not be confirmed on the basis of the gathered material. Obviously, other linguistic and non-linguistic factors such as attitudes towards the test languages (Gooskens and Heuven 2020) can influence the LID performance and should be included in the evaluation of language identification results.

3.5 Conclusions

This study had shed light on the question of involvement of native speakers without training in perceptual identification of linguistic origin when exposed to highly limited information. Work with so-called under-represented languages and vernaculars can be advanced by applying methods which combine human- and machine-based LID, especially when the available data are limited, or the training baseline is not sufficient. Since cues that are important for humans are significantly different from those for machines, it is obvious that the LID methodology should be not only source-specific but also recognizer-dependent, with respect to the domain of the training data. Even if the signal distortion reaches extreme levels, as it was in this study, in which semantic, lexical, and syntactic cues were not available, the accurate identification of the origin of a speaker was still possible by lay listeners. Therefore, the opinions of native speakers in the LADO framework should not be neglected. This study provides a clear argument for the involvement of native speakers in LADO / LOID procedures. Nevertheless, the improvement of origin verification tests using phonetic analysis remains uncontroversial. It appears that highly limited signals can cause an attention shift towards typically less relevant features in spoken language perception such as vowel quality in the spectral and temporal domains. These findings should be considered in the procedure of LADO / LOID tests, as well as in forensic investigations. It appears that speakers of closely related languages can successfully identify the linguistic origin of a user of another Slavic language, which contributes to the debate on involvement of non-experts in the LADO procedures. It also demonstrates the impressive human capability to identify the origin of a speaker with exposure to even highly limited acoustic information.

3.6 Data availability

The experimental results and scripts are publicly available in the following (https://osf.io/b4jfg/) Open Science Framework repository.

Chapter 4

Surprisal and Articulatory Gestures

The following chapter attempts to investigate the relationship between predictability and reduction on a diphone level. Since the information theoretic notion of surprisal moderates linguistic encoding, in this study such a relation will be tested in acoustic and articulatory domains. Therefore, this chapter extends the perspective by introducing a production study which involves a diphone as a unit of analyses and, therefore, attempts to shed light on the impact of diphone surprisal on a highly constraint consonant-to-vowel transitions.

The results of this experiment were presented at the ITG 2021 Conference on Speech Communication and published in the conference proceedings (Kudera et al. 2021b).

4.1 Introduction

The relation between information density and phonetic encoding is reciprocal: the more surprising the information, the more prominent the encoding. Speech production patterns on the acoustic and articulatory planes are influenced by information structure and vary depending on the distribution of information conveyed by the message (Jurafsky et al. 2001; Turk 2010). This phenomenon finds its explanation in efficient communication principles, where easily predictable and contextually enhanced elements are prone to reductions, higher degrees of coarticulation, or even omission. On the other hand, more unexpected focus with less predictable content often requires acoustic prominence, emphasis, or even reduplication (Aylett and Turk 2004; Aylett and Turk 2006).

In this study, the relation between predictability and reduction in Polish CV groups on the articulatory and acoustic levels was addressed. The information-theoretic measure of contextual unpredictability, i.e., surprisal was related with coarticulation strength. An attempt to understand how the predictability of vocalic segments corresponds with articulatory gestures and CV transitions in meaningless diphones was made. Since the influence of contextual probability on speech production has already been discovered on the morphophonological level (Shaw et al. 2014; Shaw and Kawahara 2019), the focus on such a highly constrained unit requires further justification in larger kinematic corpora, such as EMA-AME (Ji et al. 2014), MOCHA-TIMIT (Wrench 1999), DoubleTalk (Scobbie et al. 2013) and MNGU0 (Richmond et al. 2011). To address the counter-intuitive question of the 'informativity of meaningless', this investigation tries to determine whether speech production is inherently constrained by the high surprisal effect, even with the absence of semantics on the diphone level, which results in a low degree of coarticulation; or, conversely, whether the effect of articulatory and acoustic prominence given to disambiguate highly-predictable, low-surprisal sequences occurs and results in reverse patterns of coarticulation resistance. In other words, is it assumed that certain units with various information load and contextually determined frequency of occurrence require different processing effort, which results in distinctive phonetic encoding.

Therefore, this study concentrates on the diphone level to examine the role of surprisal in the highly limited and phonetically constrained environment of Polish CV sequences. In line with the Probabilistic Reduction Hypothesis (Lieberman 1963; Jurafsky et al. 2001; Turk 2010), this investigation aims to find a relation between segment unexpectedness and its acoustic and articulatory encoding, even in an experimental setup which is highly divergent from the natural speech environment.

4.1.1 Aims and hypothesis

The main goal of this experiment is to test the effect of surprisal on the articulatory gestures of Polish CV groups, composed of voiceless velar plosives and monophthongs, using acoustic and kinematic data. This pilot study aims to present the influence of surprisal on articulatory gestures and F2 dynamics with a focus on consonant-to-vowel transitions.

It is assumed that higher diphone surprisal values correspond to lower coarticulation strengths, whereas contextually predictable segments with lower surprisal values are produced with higher degrees of coarticulation. This premise relates to the applied measures in the dependency relation, that is: the lower the surprisal, the steeper the LE regression slopes of F2 measured at vowel onset and midpoint, and vice versa.

4.1.2 Related work

The relation of information density and phonetic encoding was previously investigated by (Jurafsky et al. 2001; Aylett and Turk 2004) who concluded that information-theoretic factors have an influence on phonetic encoding on various levels of representation, from segmental to suprasegmental (Aylett and Turk 2006). Similarly to this study, the anticipatory coarticulation in CV syllables was investigated by (Tabain 2000; Lee 2017) with the application of the EMA methodology. While the acoustic and kinematic data do not always correspond to each other in a linear fashion (Stevens 1989; Löfqvist 1999), many previous studies justify the approach of combining the articulatory and acoustics measures in the analyses of articulatory gestures (Fowler 1994; Recasens and Espinosa 2009). Even though in this investigation only one subject was recorded, a survey of 905 EMA studies (Rebernik et al. 2021) reports that one-participant experiments, perhaps due to fatigue of data collection, are not uncommon.

The relation between F2 and articulatory gesture on a horizontal plane can be captured by Locus Equations. Since the introduction of LEs in the 60s (Lindblom 1963), the method of presenting formant transitions on the basis of two time stamps (in onset and temporal midpoint) has been widely applied in estimating vowel reduction (Krull 1989), categorizing a place of occlusion (Modarresi et al. 2005; Sussman et al. 1991), measuring a degree of coarticulation (Löfqvist 1999; Fruchter and Sussman 1997), and expressing the articulatory synergy (Iskarous et al. 2010). More recently, such an approach has provided an insight into the articulation of CV groups in conjunction with articulographic data (Tabain 2000; Lee 2017).

The introduced notion of surprisal quantifies the contextual unit probability on various linguistic levels. With respect to the focus of this study, it is referred to the unexpectedness of CV sequences on the basis of data extracted from spoken corpora, similarly to (Shaw and Kawahara 2019). The influence of surprisal on the articulatory patterns in CV contexts can be treated as a modifier of motor fluency, which was already pointed out by (Tilsen 2014; Shaw and Kawahara 2019). In parallel to coarticulation understood as a gestural and spectral overlap of adjacent segments, a related concept, 'coarticulation resistance', was introduced by (Recasens and Espinosa 2009) and defined as the degree of articulatory variability of consonants or vowels as a function of their phonetic context. As pointed out by (Perillo et al. 2017), the correspondence of LE slopes with coarticulation resistance can be explained by the DAC (Degree of Articulatory Constraint) model (Recasens et al. 1997). By analogy to the above-mentioned studies, this chapter aims to specify contextual constraints, and focuses on the predictability of diphone segments by implementing a measure of surprisal.

4.2 Material and measurements

The CV sequences consisted of a voiceless velar stop and one from the set of Polish monophthongs: /k/ + /a, ɛ, i, ɔ, u, ɨ/. None of the monophthongs in Polish have a phonological length distinction, so the influence of segment duration was excluded from the potential set of distractors. To account for the influence of VOT (voice onset time), only one stop was tested due to the potential uncontrolled effect caused by obstruents with various long-lags, which show a tendency for steeper LE slopes when F2 onset coincides with the first glottal pulse (Engstrand and Lindblom 1997). The CV sequences were clustered into three groups depending on the surprisal. The surprisal values were calculated on the basis of the conditional probabilities of the above-mentioned vowels following the voiceless velar stop in the spoken NKJP (The National Corpus of Polish) database (Pęzik 2012) using the formula $-log_2 P(k|V)$, in which V corresponds to one of the six Polish monophthongs that could follow the consonant /k/. The result of the negative log (base 2) transformation on the conditional probability yields a measure of the unexpectedness of a vowel token in a CV context, rendered in bits. Then, the CV sequences were clustered into three surprisal-based groups. The high surprisal group was comprised of /kɛ/ and /kɨ/, the mid-surprisal group contained /ki/ and /kɔ/; and the low surprisal sequences were/ku/ and /ka/.

The recording sessions consisted of three parts: (1) an accommodation stage to allow the speaker to get used to speaking with the EMA electrodes and wires (Dromey et al. 2018), then (2) self-paced uttering of vowels in isolation, and (3) finally the articulation of the CV groups. To account for the influence of speech rate on coarticulation patterns (Bakran and Mildner 1995), the subject was asked to maintain a comfortable tempo and to utter the sequences as effortlessly as possible. To cross-check the calibration procedure, a simple real-time orientation was conducted to ensure that reference (6 DOF) sensors show a relatively large Y value, and that the tongue tip (TT) sensor had a larger X value than the tongue back (TB) sensor. After the recording stage, the V and CV sequences were automatically segmented with minor manual boundary corrections, and the Euclidean distance (ED), that is, the square root of summed squared differences between the TB coil in the V onset and the midpoint position, was calculated.

4.2.1 Acoustic measurements

Acoustic data were collected simultaneously with the kinematic data and synchronized. The mono sound was recorded with a sampling rate of 48 kHz to uncompressed .wav format. The F2 data was extracted using the LPC Burg algorithm with window length = $0.025s$; maximum F2 threshold = 5000 Hz; dynamic range = 30 dB and pre-emphasis filter from 50 Hz. The F2 values from vowel onset and midpoint were then used to derive the LEs for the respective CV groups. The LEs were represented as linear slope–intercept regressions of the F2 onset frequency transition to the F2 frequency in the middle of the vowels. After the F2 values were extracted, the LEs were computed to define the degree of coarticulation in the CV groups, according to the equation $F2_o = k \times F2_m + c$, in which $F2_o$ refers to second formant value at the vowel onset; $F2_m$ = F2 at the vowel midpoint; k = slope, which indicates the degree of change of F2 across two measuring points; and c = intercept. In order to obtain and extract these data, two 12-minute long recording sessions were conducted. The data from the first recording session were discarded due to technical difficulties and another session was recorded.

4.2.2 Articulatory measurements

Articulatory data were gathered by means of NDI Wave Articulograph with a 100 Hz sampling rate. Before the recording, the autocalibration was run to meet the standard accuracy of < 0.5 mm in the near (300 mm^3) acquisition field (Berry 2011). The naïve subject was a male native speaker of standard Polish who reported no hearing, speech, or language disorders. In total, 649 vowels in isolation and 693 vowels proceeded by voiceless velar stops were recorded from this speaker. To capture the articulatory gesture, three sensors were attached to the midsagittal plane of the tongue. The tongue tip (TT) sensor was placed 1 cm from the anatomical apex. The tongue back (TB) sensor was placed at a comfortable distance of 4 cm from the TT, and the tongue middle (TM) sensor was attached at the midpoint between the TT and TB, that is, 2 cm behind the TT coil. The two-channel six degrees of freedom (6 DOF) reference coils were attached to the left mastoid bone and nasion for head and skin movement normalization. The sensors were covered with liquid latex beforehand and glued to the dried tongue with nontoxic dental adhesive. One sensor was placed at the lower incisors for tracking the movements of the jaw. The occlusal plane was recorded with coils attached to the protractor. The palatal probe was done before the recording session and after the recommended accommodation stage with 6D palate trace sensor. Data from one recording session was discarded due to sensor displacement. Afterwards, the tongue was dried again for the coil attachment and the entire session was repeated to ensure the measurement and reference consistency.

4.3 Results

Less than 1% of the kinematic data, including sensor position and rotation in quaternion 4-D unit vector, were discarded due to missing sensor input. The corresponding acoustic data points were discarded from the analysis. The descriptive statistics, along with the Kolmogorov-Smirnov test were calculated for the dependent variables with the FSA package (Ogle et al. 2021) in R (R Core Team 2020). Then, the Kruskal-Wallis test and the post-hoc Dunn's test with Holm's correction for multiple comparisons were performed.

The results of the Kolmogorov-Smirnov test showed significant deviance from normal distribution of five tested variables across the surprisal groups, that is: in the low surprisal cluster (LE: $D = 0.13, p <$

0.001, ED: $D = 0.11, p < 0.001$); the mid-surprisal (LE: $D = 0.26, p < 0.001$, ED: $D = 0.14, p < 0.001$); and the high surprisal (LE: $D = 0.12, p < 0.001$, ED: $D = 0.04, p = 0.200$). Therefore, the Kruskal-Wallis test was conducted to assess the differences in ED and LE between the surprisal groups.

4.3.1 Acoustic domain

A steep slope across tokens indicates maximal coarticulation and corresponds to small differences between the F2 at onset and midpoint with a high overlap of the segments. In contrast, a shallow slope corresponds to a low coarticulation effect and minimal overlap of the adjacent segments.

On the basis of the gathered acoustic data, a strong effect of the surprisal group on the LE slopes was discovered ($H(2) = 138.71, p < 0.001, \epsilon^2 = 0.20$). The post-hoc tests showed higher coarticulation for the middle surprisal groups, compared with low surprisal ($Z = 11.13, p < 0.001$) as well as high surprisal clusters ($Z = -8.87, p < 0.001$). Additionally, the low surprisal group had steeper slopes than the high surprisal groups ($Z = -2.23, p = 0.020$), see Figure 4.1 4.1 where low LE correspond to low coarticulation.

4.3.2 Kinematic domain

The Euclidean distance between the position of the TB electrode measured at the two timestamps reflects the articulation resistance in relation: the larger the distance, the lower the degree of coarticulation. The analyzed set of kinematic data showed a similar correspondence to the one discovered on the basis of the acoustic data but with moderate magnitude ($H(2) = 83.268, p < 0.001, \epsilon^2 = 0.12$). The post-hoc test showed a higher coarticulation for the middle surprisal groups compared with low surprisal ($Z = -8.05, p < 0.001$), as well as high surprisal groups ($Z = -7.67, p < 0.001$). In the kinematic dataset, no significant differences were observed between the surprisal boundary groups ($Z = 0.36, p < 0.717$), see Figure 4.2, where High ED correspond to low coarticulation.

4.4 Discussion

This pilot study aimed to estimate the degree of coarticulation on the level of the diphone CV transition. It was anticipated to discover an

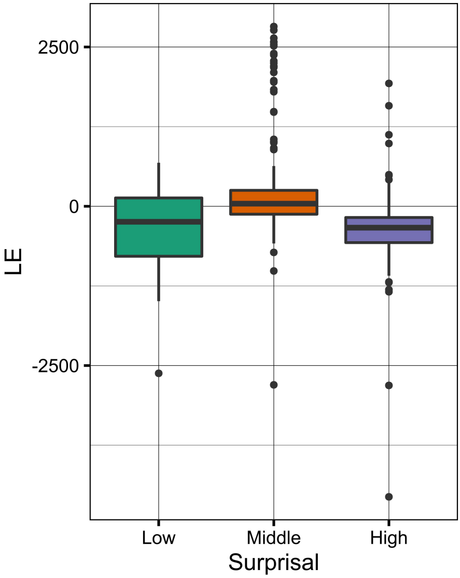

Figure 4.1: LE slopes across three groups of surprisal. Low LE values correspond to low coarticulation effect.

effect of the unexpectedness of a segment on the articulatory gesture of a highly constrained cluster. Since the coarticulation strength cannot be inferred from the formant values only, a hybrid methodology was applied involving a combination of spectral measures with the kinematic data. On the basis of measurements from two different domains, it was concluded that the degree of coarticulation of the Polish CV groups is the highest in the middle surprisal cluster, whereas the low and high surprisal groups are characterized by higher coarticulation resistance and lower spectral overlap. Hence, the initial expectations were confirmed only in part.

4.4.1 Discrepancies across domains

Possibly due to limited size of the recorded data, the discrepancies between the acoustic signal and its kinematic source were observed. Even though a linear correspondence was not discovered, the patterns are consistent across the domains, which complements the argument that the

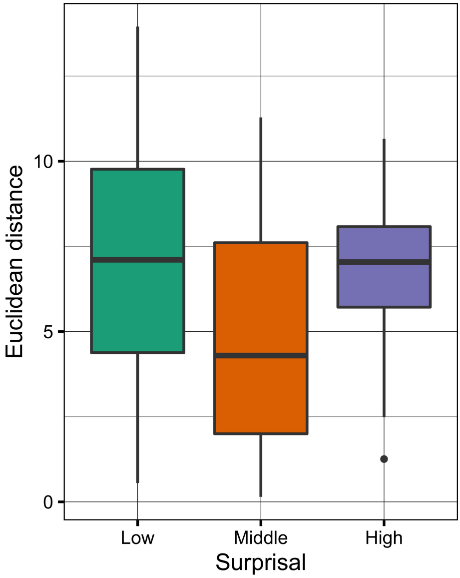

Figure 4.2: ED [in mm] across three groups of surprisal. High ED values correspond to low coarticulation effect.

LE measure does not necessarily reflect the motor functioning of the articulator (Fowler 1994; Löfqvist 1999). Furthermore, formants are influenced by too many factors to obtain a clear TB sensor position aligned with F2 fluctuation. Hence, one of the reasons for why the tendencies across the domains are approximated rather than ideal can be traced to the positioning of the TB coil on the tongue. It is assumed that more posterior placement might contribute to higher correlation between the acoustic and kinematic data measured on the basis of the second formant value and the TB sensor position.

4.4.2 Cluster synchrony effect

In the spectral domain, the minor differences between the F2 measures also seem to support the CV synchrony hypothesis. Furthermore, it should be noticed that the experimental setup itself can cause hyperarticulated speech (Moon and Lindblom 1994), which in consequence can lead to variations in articulatory gestures (Fougeron and Keating 1997).

The experimental design with monotonous repetitions can also influence the degree of anticipatory coarticulation due to the increased effort of speaking in lab conditions with a bundle of wires. The diversity of the recording data can also strengthen the outcomes because of the possible influence of the competition effect of habituation, which refers to the effortless pattern of articulation in EMA. Such an experimental setup might involve gestures that diverge from the natural speech.

4.5 Conclusions

It is concluded that coarticulation measured on the diphone transition is a consequence of a twofold surprisal effect. Firstly, regarding the small degree of coarticulation in the low surprisal groups, it can be assumed that the behavioural pattern contributes to the need for segment disambiguation in the absence of the supportive context of surrounding meaningful elements. Secondly, the low coarticulation effect observed in the high surprisal sequences adheres to an underlying pattern of giving prominence in articulation to segmental concatenations with low predictability. In support of both explanations, the middle surprisal group showed the highest degree of coarticulation across the domains, hence neither the need for disambiguation nor the prominence effect was present in frequent phonotactic sequences. The observed effect can also be attributed to motor practice, which suggests an improvement of the articulatory skills with frequency of use (Tomaschek et al. 2013; Tomaschek et al. 2021). Motor practice is associated with a stronger overlap of two adjacent gestures and increases with the frequency of sequence occurrence. Increasing experience in motor practice can also lead to enhanced kinematic movements and more extreme articulation (Tomaschek et al. 2018).

To further extend this study, a larger pool of participants would be desired. The increase of the sampling rate to 400 Hz can also bring higher resolution in the motion capture data. Regarding the information-theoretical approach, the further extension of this pilot study might involve, apart from surprisal, entropy as an index of competition in vowel selection.

Even though the aforementioned methodological issues should be addressed in further studies on the effect of surprisal on articulatory gestures, the interplay of information-theoretic and phonetic measures was discovered. It was concluded that phonotactic predictability can induce changes to speech production patterns even on the highly constrained diphone level. The results appear consistent with previous studies investigating the relation between coarticulation strength and contextual predictability. Interestingly, it was discovered that the relation between surprisal and phonetic encoding occurs even on the segmental level and affects meaningless syllables uttered in isolation. This evidence provides an argument for the inherent nature of a surprisal component in language production and suggests that the phonotactic unexpectedness of a segment influences motor fluency.

4.6 Data availability

The datasets are publicly available in the following (https://osf.io/va3ge/) Open Science Framework repository.

Chapter 5

Cross-lingual Priming

Following the structuralist narrative, this chapter expands the scope from phonetic to lexical units and attempts to investigate comprehension phenomena in lexical decision task conducted in short-term priming framework. The applied lexical stimuli were categorized into two conditions, namely, associative and phonetic priming.

The results of this study were presented at the Interspeech 2021 conference and published in the conference proceedings (Kudera et al. 2021a).

5.1 Introduction

Priming is a general property of human cognition that refers to a behavioral response after a sequence of stimuli which are related to each other within or across modalities. A recently experienced stimulus, i.e., a prime (Token 1), influences the way one responds to a target, that is, the later stimulus (Token 2). This is true not only in lexical access tasks, but also across various sensory modalities such as auditory, visual, and olfactory. The relation between the stimuli is conventionally reflected by means of the response time (RT) measured in behavioral tasks. This relation can cause a facilitation effect, namely shorter latency, or an inhibition effect, whereby the response to a target input is delayed due to a distant, unclear, or undiscovered relation between the perceived stimuli.

In terms of human language processing, the perceptual priming effect depends on lexical, syntactic, morphological, and phonetic relations between primes and targets. The relative frequency of occurrence of stimuli, as well as the knowledge of the person experiencing the sequence, also plays a role, and hence influences the reaction to the stimuli in so-called positive or negative priming. In a cross-linguistic context, the size of these effects corresponds with phonetic similarity and, more broadly, typological relatedness between the languages of primes and targets (Jo-ordens and Becker 1997; Nicol 1996; Luce et al. 2000). Studies on priming in the auditory modality have shown that correspondence between stimuli is especially relevant on the phonetic and phonological levels of speech processing (Slowiaczek and Pisoni 1986; Sereno and Jongman 1992; Gor 2018); however, most investigations thus far have been conducted on monolingual datasets, rarely taking a cross-lingual perspective (Lalor and Kirsner 2000; Duyck 2005; Dijkstra et al. 2010).

To address this lacuna, this study proposes new methods for quantifying relatedness and similarity between cognates and near-homophone tokens in a multilingual setting. The information-theoretic notion of surprisal (Shannon 1948), which measures the (un)expectedness of an outcome, is introduced and correlated with latency measures obtained from a lexical decision task in a short-term priming framework involving four Slavic languages: Bulgarian, Czech, Polish, and Russian.

5.1.1 Aims and hypotheses

The primary goal of this chapter is to present measures of cross-lingual speech comprehension based on phonetic distance and surprisal, and to validate them in a behavioral priming task. By grounding the perceptual study in information-theoretic methodology, the following assumptions will be addressed: (1) facilitating priming is present as an effect of exposure to multilingual tokens from closely-related languages; (2) the priming type (cognate vs. phonetic) as well as the language of stimuli influence latency in a lexical decision paradigm; (3) behavioral reaction, measured as response time, depends not only on the similarity between successive primes and targets but also on their (un)expectedness. Thus, the following hypotheses were tested: phonetic proximity of closely-related Slavic languages shortens the latency, regardless of the language of the prime; cognates shorten response times to a greater extent than do near-homophones in a multilingual experiment; and the information-theoretic approach, by introducing a unit of (un)expectedness, outperforms canonical measures of similarity between prime and target.

5.1.2 Related work

Previous studies on priming effects involving stimuli from several languages have shown that these effects depend on various linguistic levels. Semantic and etymological relatedness between tokens cause a cognate facilitation effect (Costa et al. 2000; Peeters et al. 2013). Furthermore, discussions on latencies in lexical decision tasks have supported theories of associative activation during the process of searching for unique mental representations of stimuli. The strength of the facilitation effect can be asymmetric and reflect the subjects' dominant language, which might correlate with the size of one's lexicon or, as suggested by (Dijkstra 2005), differences in conceptual representations between a dominant language and an L2. In a multilingual scenario, RT was also applied as a direct measure of similarity across related languages, under the assumption that shorter response times reflect better intelligibility of subjects' L2 (Gooskens 2013).

5.2 Methods

The study consisted of a lexical decision task in a priming paradigm using spoken stimuli, which were either cognate pairs or semantically-distant near-homophones. In the pre-test stage of the experiment, the relation between primes and targets was measured using phonetic distance and word adaptation surprisal. The degree of phonetic similarity was measured as a feature-based phonetic distance of aligned segments and their sequence within words. In the experiment, native speakers of four Slavic languages listened to primes (Token 1) in the three non-native languages and were asked to decide whether the targets (Token 2) were truly words in their native language. The targets in all cognate and near-homophone pairs were in the participants' L1, but were interspersed with filler word pairs (in a ratio of 4:1), in which the targets were not in the participants' L1 and were both phonetically and semantically distant from the primes. Each participant was exposed to 156 prime-target pairs. The subjects were instructed that their decision should be as accurate and fast as possible. Participants were paid for completing the study and additionally motivated by a bonus payment awarded for the best performance in each language group. On-screen visual feedback was presented to the participants in the form of a real-time plot of their accuracy and amount of time taken per task. The set of fillers was discarded from further analysis. The participants were presented with randomized pairs of primes and targets, and their decisions and response times were recorded.

5.2.1 Stimuli

The stimuli consisted of read-speech samples of true cognates and near-homophones identified by phonetic distances in Bulgarian, Czech, Polish, and Russian. The near-homophone pairs were automatically identified from transcribed wordlists on the basis of small pairwise phonetic distances. The cognate stimuli were used in the associative priming condition, whereas the near-homophone stimuli were used in the phonetic priming condition. The tokens, both primes and targets, controlled for frequency, were extracted from recordings of female native speakers of each language in self-paced reading sessions of token lists. The readings were recorded in an acoustically controlled environment with a 48 kHz sampling rate to uncompressed format. Two recording sessions were conducted for each native speaker.

The cross-lingual phonetic distance between primes and targets was calculated as the weighted sum of three component scores: dissimilarity of consonantal segments (0.5), dissimilarity of vocalic segments (0.3), and difference in syllable structure (0.2). Segments in word pairs were first aligned automatically using the Needleman-Wunsch algorithm (Needleman and Wunsch 1970), with alignment costs based on segment pairs' differences in distinctive features and sonority. The cosine similarity of distinctive feature vectors was then taken for each pair of aligned segments, with gap sequences (a segment aligned to nothing) and alignments of vowels to non-glide consonants receiving a similarity score of zero.

The overall similarity of consonant segments was calculated as the mean of cosine similarities of aligned consonants, and the similarity of vocalic segments was likewise calculated as the mean of the cosine similarities of the aligned vowels. Each of these similarity values was then subtracted from the maximum similarity of 1.0 to yield a measure of dissimilarity.

The difference in syllable structure was quantified as the length-normalized Levenshtein distance of each word pair encoded as sequences of 'C' (consonant) and 'V' (vowel). For example, the Polish word *chłodny*

/xwɔdnɨ/ 'cold' was rendered as CCVCCV, whereas its Bulgarian counterpart *хладен* /xɫadɛn/ was rendered as CCVCVC. Hence, the distance between inter-lingual homophones exhibiting identical features and syllable structure was equal to zero, and the higher the phonetic distance, the more phonetically distinctive the primes and targets. A maximum phonetic distance of 1.0 would only be reached in the case of zero aligned segments and no overlap in syllable structure (e.g., a word consisting of a single consonant and a word consisting of a single vowel).

The (un)expectedness of the phonetic form of the targets (Y) given their primes (X) was quantified by means of surprisal, measured in bits, given in the following equation:

$$surprisal(Y|X) = -log_2 P(Y|X)$$

The surprisal measures were calculated for pairs of stimuli using a method based on word adaptation surprisal (WAS), which has been presented as a model of inter-comprehension among Slavic languages in orthographic texts (Stenger et al. 2017). As an adaptation to the spoken modality, word adaptation surprisal was calculated on the phonetic level using IPA transcriptions of stimuli. Probabilities of inter-lingual phonetic correspondences were extracted from 1030 automatically aligned pairs of true cognates in Bulgarian, Czech, Polish, and Russian. These correspondence probabilities were then used to calculate the length-normalized phonetic word-adaptation surprisal between prime and target word pairs used in this study, according to following equation:

$$WAS = \frac{1}{n}\sum_{i=1}^{n} -log_2 P(L1_i|L2_i)$$

where: $L1_i = i^{th}$ phone in native (decoder) language and $L2_i = i^{th}$ phone in foreign (stimulus) language. Hence, surprisal quantifies the informativity of cross-linguistic phoneme correspondences for aligned pairs of primes and targets.

5.2.2 Participants

In total, 200 participants (50 native speakers of each tested language) recruited via an online crowd-sourcing platform took part in the experiment. On the basis of a pre-test questionnaire, participants with diagnosed hearing disorders or a background in Slavistics were excluded

Table 5.1: Lexical decision task results - descriptive statistics

Variable	*M*	*Me*	*SD*	*Skew*	*Kurt*	*Min*	*Max*	*D*	*p*
RT	902.1	888.68	227.15	0.14	0.07	151.63	1499	.03	< 0.001
Phon. dist.	0.08	0.07	0.06	0.36	-1.30	0	0.20	.16	< 0.001
Surprisal	2.34	1.64	1.90	1.25	1.04	0.05	8.91	.18	< 0.001

from the analysis (less than 1% of participants). The analyzed dataset was then supplemented with additional experimental sessions to ensure equally balanced groups of participants for each language.

5.2.3 Data analysis

Only correct answers, not including responses to fillers, were considered in the analysis. The RT range (150 ms $<$ RT $<$ 1500 ms) was altered in accordance with studies measuring monolingual priming effects (Perea and Rosa 2002). The typical floor was set to filter out accidental responses, whereas the ceiling defined the threshold of the short-term priming effect. Each prime–target pair was repeated three times over the course of the experimental session, and a participant's RT for a particular prime–target pair was calculated as the mean of their three RTs for this pair. Basic descriptive statistics were calculated with the Kolmogorov-Smirnov test. A three-way ANOVA was applied for independent samples, and a Pearson correlation and moderation analysis was carried out to validate phonetic distance and surprisal – the methods proposed to quantify the cross-lingual correspondences.

5.3 Results

The Kolmogorov-Smirnov test showed that the distribution of mean results in both the pre-test and the post-test was significantly different from the normal distribution. However, both skewness and kurtosis were smaller than the absolute value of 2, so a parametric test could be performed. The descriptive statistics are presented in Table 5.1. In order to verify the differences in response times depending on priming type (cognate vs. phonetic) across languages, a three-way ANOVA for independent samples was performed. A statistically significant main effect for priming type was found $F(1, 4344) = 95.43; p < .001; \eta^2 = .02$.

Hence, on the basis of the gathered data, the cognate facilitation effect was confirmed and response time was higher in the phonetic priming type. The language of primes did not appear to affect the results. The Token 1 language main effect was not statistically significant, $F(3, 4344) = 0.59; p < .619; \eta^2 = 0$. Therefore, no post-hoc analysis was performed.

5.3.1 The Slavic L1 effect

The four groups of subjects did not differ significantly in their measured RTs. Furthermore, no statistically significant interaction effect between priming type and Token 2 language was found: $F(3, 4344) = 0.49; p = .690; \eta^2 = 0$. On the other hand, a statistically significant Token 2 language effect was found: $F(1, 4344) = 122.68; p < .001; \eta^2 = .08$. A post-hoc analysis by means of Sidak tests was then performed. Response times in Polish were significantly lower than in the other three languages ($p < .001$). Response times in Czech were also significantly lower than in Bulgarian and Russian ($p < .001$). Bulgarian and Russian response times were not significantly different from one another. Response times across the four language groups are plotted in Figure 5.1.

Furthermore, a statistically significant interaction effect between priming type and Token 1 language type was found; $F(3, 4344) = 3.82; p = .01; \eta^2 = .003$. Therefore, a simple effect analysis was obligatorily performed. A statistically significant effect of the language of Token 1 was found, but only in the phonetic priming condition, $F(3, 4344) = 11.60; p < .001; \eta^2 = .008$. In total, three statistically significant differences in the post-hoc analysis were discovered. In the cognate priming condition, the Token 1 language effect was not statistically significant, $F(3, 4344)S = 2.13; p = .094; \eta^2 = .001$. On the other hand, a significant priming type effect in all languages was discovered - Polish: $F(3, 4344) = 67; p < .001; \eta^2 = .015$; Czech: $F(3, 4344) = 19.29; p < .001; \eta^2 = .004$; Bulgarian: $F(3, 4344) = 16.45; p < .001; \eta^2 = .004$ and Russian: $F(3, 4344) = 11.40; p = .001; \eta^2 = .003$. In line with the cognate facilitation effect, the RT values were higher in the phonetic priming type.

In order to verify the asymmetric character of relation among the language groups, the post-hoc analysis was performed. Responses to Polish primes given by Czech native speakers were significantly lower compared to Bulgarian ($p = .018$) and Russian ($p = .001$) language groups.

Table 5.2: RTs across language groups and conditions

		Cognate		Phonetic		Both cond.	
Token 1	Token 2	*M*	*SD*	*M*	*SD*	*M*	*SD*
Polish	Czech	775.33	114.15	924.54	192.78	849.94	185.33
	Bulgarian	854.67	237.63	960.60	220.35	907.23	235.09
	Russian	866.15	233.90	987.39	233.58	924.40	241.26
	all	851.82	229.98	967.53	223.60	908.55	234.05
Czech	Polish	699.72	165.56	761.70	171.78	730.16	170.99
	Bulgarian	910.93	212.15	976.88	193.99	943.76	205.83
	Russian	902.03	215.57	955.61	227.11	928.19	222.68
	all	881.28	219.02	942.47	215.35	911.47	219.29
Bulgarian	Polish	681.86	189.89	775.90	194.97	727.26	196.03
	Czech	794.02	160.21	875.07	154.02	831.80	161.86
	Russian	927.70	223.74	973.81	225.30	950.25	225.46
	all	852.83	230.87	916.22	224.93	883.52	230.07
Russian	Polish	742.40	202.03	783.44	190.29	763.04	196.70
	Czech	791.58	158.49	859.25	174.23	825.15	169.25
	Bulgarian	909.81	215.68	981.21	229.54	943.61	225.02
	all	865.53	217.47	927.25	229.87	895.26	225.53
all	Polish	707.19	186.26	773.73	185.48	739.94	188.65
	Czech	787.04	153.90	886.73	176.46	835.66	172.44
	Bulgarian	892.32	223.18	972.82	214.40	931.71	222.54
	Russian	898.20	225.73	972.20	228.76	934.15	230.12
	all	864.07	224.34	941.95	223.26	902.10	227.15

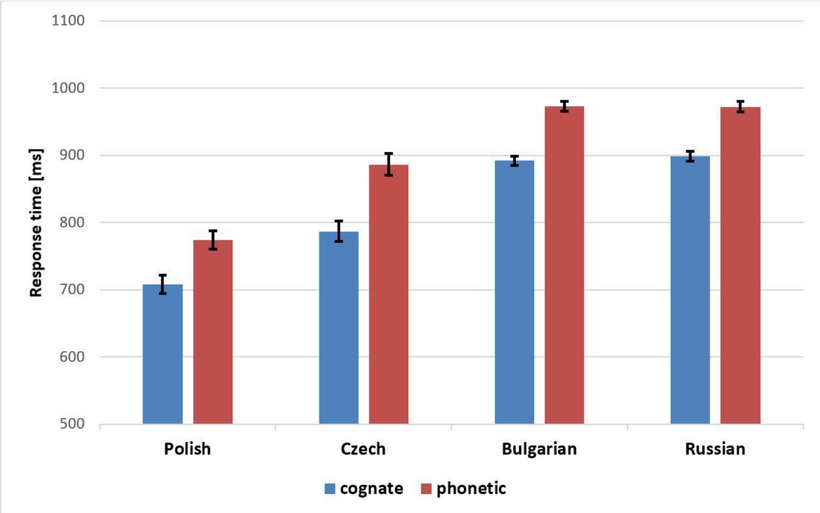

Figure 5.1: RTs in cognate and phonetic priming according to participants' L1

5.3.2 Moderation analysis

In the following step, a Pearson correlation analysis was performed to find out whether phonetic distance and surprisal measures influence response times. A statistically significant correlation was discovered in both analyses. The phonetic distance measure was positively correlated with latency ($r = .190; p < .001$). As expected, larger phonetic distances were associated with longer response times. The measure of (un)expectedness via phonetic word adaptation surprisal was also positively correlated with RTs ($r = .219; p < .001$). Again, the higher the surprisal values – in other words, the higher the degree of unexpectedness – the more time subjects needed for their responses.

Subsequently, a moderation analysis was conducted to validate the quantification measures in the behavioral test. The phonetic distance and surprisal scores were clustered into three groups to verify their influence on the RT variable. It appeared that the latency moderation effect of phonetic distance is strong for Polish ($B = 946.34, SE = 90.42, t = 10.47, p < .001$) and Czech ($B = 750.38, SE = 91.03, t = 8.24, p < .001$), whereas no significant effect was found for Bulgarian or Russian. The results are plotted in Figure 5.3.

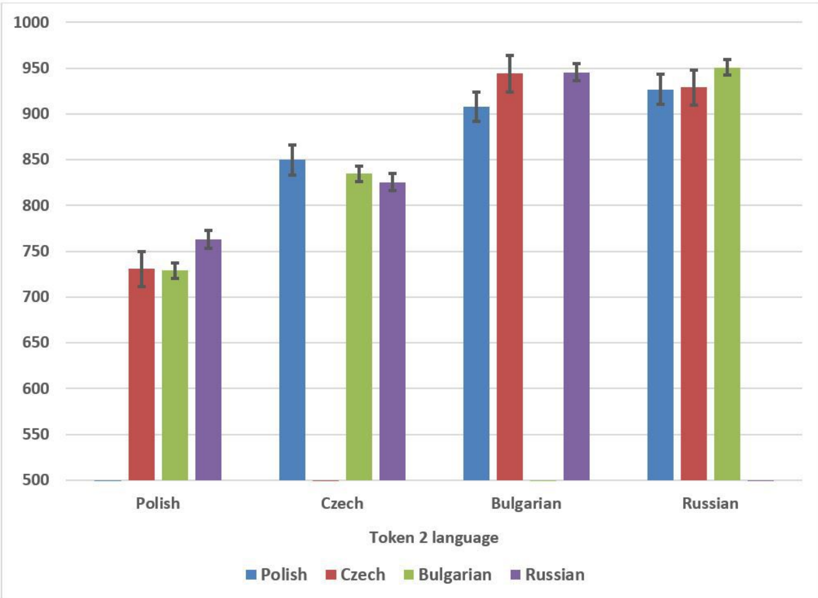

Figure 5.2: RTs in both conditions (phonetic and associative) across language groups

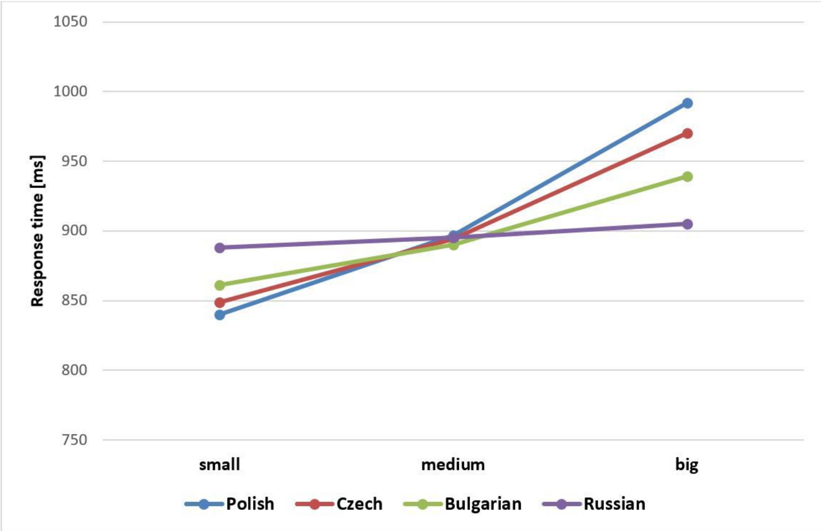

Figure 5.3: RTs across three levels of phonetic distance for all tested languages

The information theoretical measure of surprisal significantly moderates the RT in the data obtained from this study. It outperforms the phonetic distance measure and applies to all four tested languages: Polish ($B = 33.58, SE = 3.18, t = 10.55, p < .001$), Czech ($B = 17.54, SE = 2.79, t = 6.28, p < .001$), Bulgarian ($B = 32.53, SE = 5.61, t = 5.79, p < .001$), and Russian ($B = 28.24, SE = 3.96, t = 7.14, p < .001$). The results are plotted in Figure 5.4.

5.4 Discussion

In this short-term priming study, two methods of quantifying phonetic similarities between cognates and near-homophones were tested in a multilingual lexical decision task experiment. One based on the degree of similarity in phonological features among corresponding phones; whereas the other, grounded in information theory, was based on the (un)expectedness of stimuli, measured in bits. These two methods were applied in order to quantify the relations among Bulgarian, Czech, Polish, and Russian targets and primes, with latency measures from the

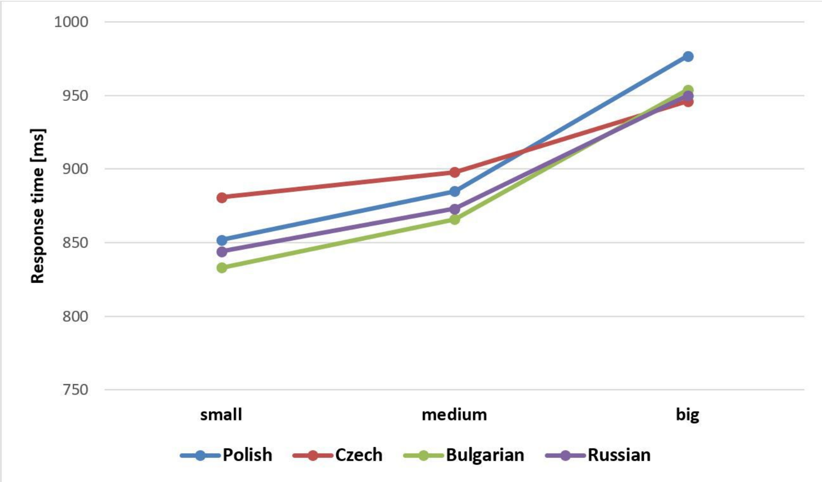

Figure 5.4: RTs across three levels of surprisal for all tested languages

behavioral lexical decision task taken as a validation reference. The outcome of this study sheds light on the cognitive processing of linguistic relatedness and phonetic similarity of spoken stimuli in closely-related Slavic languages.

5.4.1 Cross-lingual priming effect

The first hypothesis, regarding the cognate facilitation effect, was supported in the multilingual priming experiment. It seems that the proximity of the Slavic languages is a crucial factor that enables an immediate recognition of cognates and thus promotes the facilitation effect. The discrepancies in the results among the different language groups, however, suggest asymmetric intelligibility effects. This finding could be further tested on a collection of more distantly related languages with more divergent phonemic inventories.

5.4.2 Priming type effect

The second hypothesis was supported with respect to the effect of priming type. In line with expectations, cognates sharing a semantic field in the Slavic languages facilitate responses to a significantly greater extent than unrelated word pairs identified by their phonetic similarity. This outcome suggests that in lexical access tasks, a primary role should be directed towards the semantic relatedness rather than to similarity based on surface representations.

5.4.3 Phonetic distance and surprisal

Thirdly, the information-theoretic approach for presenting multilingual lexical relatedness was justified. With regard to the application of phonetic distance and (un)expectedness, the new proposals, and especially the information-theoretic notion of surprisal, seem to be valid moderators of human response times after exposure to meaningful stimuli in a closely-related language. As a parallel to the positive priming effect, the influence of (un)expectedness of targets was found: the higher the surprisal, the longer the latency. This suggests that quantification of the cognate facilitation effect should be supplemented with a surprisal component, which corresponds to the (un)expectedness of stimuli.

5.5 Conclusions

The experimental results confirmed the perception of phonetic proximity of four closely-related languages from the Slavic branch of the Indo-European family. It was reported that phonetic word adaptation surprisal and cross-lingual phonetic distance between primes and targets are moderators of latency in an auditory lexical decision task. The facilitating effect of cognate tokens clustered in the associative priming condition relates to the phenomenon of Slavic receptive multilingualism. This finding provides an argument for the primacy of associative correspondences and subjects' ability to identify the semantic relatedness of stimuli from another Slavic language. The participants' ability to immediately recognize the associative link between a non-native Slavic prime and a target in their Slavic L1 contributes to a strong intercomprehension effect, even among individuals without any formal education or training in linguistics or Slavistics. Another outcome of this study turns the focus from lexical and phonetic similarity between tokens to the contextual (un)expectedness of stimuli.

The phase of method validation, conducted on four languages, provided evidence for context-based word adaptation surprisal outperforming the classical measures based on similarity between primes and targets in a short-term priming paradigm. Whereas the latency scores obtained from this study are moderated by the (un)expectedness of tokens and explain the effect in all four languages, the phonetic distance seems to moderate RTs only in Czech and Polish tokens. The results suggest that context-based methods for establishing the relation between two meaningful words in closely related languages are better predictors of human performance than metrics established exclusively on the basis of the similarity of stimuli.

5.6 Data availability

The data are publicly available in the following (https://osf.io/jdxw9/) Open Science Framework repository.

Chapter 6

Sentence Comprehension in Visual World Paradigm

Since the previous study centered around comprehension of lexical entries, a natural extension of the perspective will introduce comprehension test focusing on a sentence level. Hence, this chapter presents the study into online comprehension of sentences in visual world paradigm.

A part of this chapter is based on the work published in proceedings of ESSV 2022 - 33rd conference on electronic speech signal processing, University of Southern Denmark, Sønderborg (Kudera et al. 2022)

6.1 Introduction

This chapter presents the results of an eye tracking experiment which aimed to test comprehension of sentences coming from closely related Slavic languages. The visual world paradigm (VWP) format allows for the measurement of language comprehension in real time through anticipatory eye movements (Kamide et al. 2003; Tanenhaus et al. 1995; Sekicki and Staudte 2018). In this study, a VWP eye tracking experiment (Duchowski 2017; Carter and Luke 2020) was employed to investigate the influence of phonetic distance and surprisal on sentence comprehension in a multilingual setting including four Slavic languages, namely, Bulgarian, Czech, Polish, and Russian.

The underlying assumption for speech processing and visual scene inspection was given by (Cooper 1974), who observed that when participants are simultaneously exposed to spoken and visual stimuli, their eye movements are synchronized with different linguistic events conveyed in the auditory modality. This observation laid the foundation for VWP and allows for examining whether participants use predicative constraints to shift their attention to the visual referent (Altmann and Kamide 1999). This study proposes an extension of such a paradigm to a multilingual setting. Therefore, with the application of the VWP method, it is possible to examine if participants turn their attention to a visual field representing the direct object of a sentence from a closely related language, perceived in the auditory modality. It is assumed that such an effect can be enhanced by similar surface forms of verbs in both languages, measured by their phonetic resemblance to one another, as well as by a degree of stimulus (un)expectedness.

6.1.1 Aims and hypotheses

The primary goal of this study was to investigate the comprehension of sentences from a closely related language in the auditory modality with a visually enhanced environment. It was assumed that information extracted at the predicate, coming from a closely related language, can be used to guide eye movements to whichever object in the visual context satisfies the restriction of a perceived predicate. Furthermore, it was hypothesized that comprehension of a sentence, measured by a gaze anticipation effect in the VWP setup, is driven by the information-theoretic notion of surprisal as well as phonetic distance between corresponding predicates in subjects' L1 and the stimulus language. Since previous studies have shown a unilateral pattern of spoken intercomprehension among speakers of the Slavic languages, a secondary aim of this work was to investigate the directionality of sentence comprehension across the language groups.

6.1.2 Related work

Eye tracking in the VWP has emerged as an important method for understanding real-time language comprehension. Previous studies using this methodology have shown a close alignment between fixations and estimates of lexical activation (Farris-Trimble and McMurray 2013). Importantly to the presented study, phonetic similarity has been shown to trigger an attention shift to a referent object (Allopenna et al. 1998). Past VWP studies on non-native spoken word recognition have shown a greater lexical competition for non-native input, but confirmed the unidirectional effect of comprehension, even when investigating less closely languages than the ones tested here, such as Dutch and English (Weber and Cutler 2004). In accordance with previous experiments, which have shown that surprisal is not determined solely by linguistic context but can be also enhanced by a visual environment (Sekicki and Staudte 2018), the presented study introduced a situated version of surprisal by merging the cross-lingual predictability of a predicate with a corresponding visual scene. The studies mentioned above demonstrate that VWP is a suitable method for investigating verb-mediated referential processing (Altmann and Kamide 1999) and therefore could be extended to a multilingual comprehension scenario.

6.2 Methods

A VWP study was designed involving four Slavic languages. A scalable webcam eye-tracker, Webgazer (Papoutsaki et al. 2016), was used for gaze estimation and the detection of visual field preferences. Subjects were exposed to auditory stimuli presented simultaneously with a visual scene. The participants were instructed to listen to the sentences and look at the pictures. Then, a pseudo-task involving answering a question in their L1 regarding their understanding of the foreign sentence was given. Subjects were informed that pictures can provide clues toward sentence comprehension, therefore they should pay attention to the objects presented on the screen. The start of the next trial was triggered after the user had recorded an answer. Head movements were unrestricted during the recording session. The experiment setup precluded the use of a chin rest, but the head pose was tracked in the background. Since syntactic constraints can influence sentence comprehension in a VWP setup (Altmann and Kamide 1999), in this experiment stimuli were composed of fixed phrases, familiar to the subjects already after the trial session. To ensure data quality, drift check intervals were presented after each visual scene (Carter and Luke 2020). In case of a lost mesh, participants were asked to adjust their head position and to recalibrate the eye tracker prior to starting the following trial. The experiment lasted around 25 minutes, although the length varied depending on subjects' ability to maintain a consistent head pose.

6.2.1 Phonetic distance

Phonetic distances between transcribed verb pairs were calculated using a weighted average of three component measures expressing phonetic dissimilarity on the word level: dissimilarity of consonantal segments (0.5), dissimilarity of vocalic segments (0.3), and dissimilarity of syllable structure (0.2) (Kudera et al. 2021a). This weighting scheme was set according to the hypothesis that consonantal segments are the most salient phonetic factor in an intercomprehension scenario, and may impact the comprehension of related word forms to a relatively greater extent than dissimilarities in the vowel space or syllable structure. Phonetic dissimilarities between individual consonant or vowel pairs were calculated from feature vectors representing each segment's phonological distinctive features. The final component quantifies the dissimilarity of the syllable structure of the two words by calculating the length-normalized Levenshtein distance (Levenshtein 1965) of the IPA strings encoded as sequences of 'C' (consonant) and 'V' (vowel) characters.

6.2.2 Surprisal

Surprisal, or self-information, is an information theoretic measure that quantifies the (un)expectedness of a particular outcome, measured in bits, inversely proportional to its probability (Shannon 1948). More specifically, surprisal is calculated as the negative logarithm (base 2) of the probability of the outcome. Thus, outcomes with higher probabilities produce lower surprisal values, and, conversely, less likely outcomes yield higher surprisal. A perfectly predictable outcome with probability of 1.0 results in a surprisal value of zero. Surprisal of verb pairs was measured according to Word Adaptation Surprisal (WAS) (Stenger et al. 2017), given in following equation, applied to aligned phonetic transcriptions in IPA:

$$WAS = -\sum_{i=1}^{n} \log_2(p(wordL1_{[i]}|wordL2_{[i]}))$$

where $p(wordL1_{[i]}|wordL2_{[i]})$ represents the probability of the $[i]^{th}$ phonetic segment of a word in a listener's native language given its aligned equivalent $wordL2_{[i]}$ in a non-native language.

This assesses the total surprisal of the sounds of one word given aligned equivalents in another word. Probabilities of phonetic correspondences between each pair of languages were extracted automatically from pairwise phonetic alignments of 434 cognate sets in the four languages. WAS has been shown to correlate with intelligibility among Slavic languages (Stenger et al. 2017) and can be interpreted here as a quantification of the processing effort required by a native speaker of one language to comprehend lexical items from a related language. To account for the differing lengths of words, the WAS values were normalized by the length of the phonetic alignment.

6.2.3 Regions of interest

Four regions of interest were predefined for each visual scene. The four pictures were presented in the corners of a screen at 150 x 150 pixels, equally distanced from the center of the screen, which depicted the agent of each carrier phrase, a girl. The girl in the picture had closed eyes to avoid the suggestion that she was pointing to any one visual field. No extraneous visual elements were included in the pictures to ensure the visual salience of the object. Clip art images depicted easily recognizable nouns of similar complexity. The images were controlled for visual complexity, quality, and size. One corresponding object was randomly placed at each visual scene. For instance, the referent object in a Polish sentence *Ona chce pić teraz kawę*, glossed as follows:

Ona	*chce*	*pić*	*teraz*	*kawę*
3SG	want-PRS:3SG	drink-INF	now	coffee.ACC:SG

'She wants to drink a coffee now'

was the only drinkable object (a cup of coffee) among the images presented, which were otherwise not objects that can be drunk or eaten. The visual field of each object in the scene was defined by its outermost contour. The trials were randomized for each recording session, and images were randomly assigned to each of the screen corners.

6.2.4 Audio stimuli

Audio stimuli consisted of fixed SVO-type sentences in four Slavic languages. The intervals from the verb offset to the onset of a filler, and from the filler offset to the object onset, were equalized across the utterances to 200 ms. The verbs and objects were separated by a filler word *now* in the relevant language. All the verbs were transitive and controlled for high collocation strength with their respective direct object. Due to the relatively flexible word order in the Slavic languages, the structure of the stimuli with a filler preceding the object sounded natural and did not violate any syntactic constraints. The audio samples were synthesized with a TTS system for each tested language. The intonation contour was manually optimized to achieve a natural sound in sentences that included two short pauses. Such a segmentation scheme also allowed for equalizing the width of the time window of the analyzed fixations. Only female voices were used in the synthesis and each sample has been verified for intelligibility and naturalness by native speakers of the tested languages.

6.2.5 Participants

In total, 100 participants (25 native speakers per each tested language) aged 19–57 (mean 28) completed the experiment. All of the participants reported using their declared L1 in everyday communication. Only data from subjects who reported no vision or hearing related difficulties were included in the analysis. Therefore, no glasses were allowed. Participants were recruited via a crowd-sourcing platform and paid for their participation. Importantly, subjects had not received previous training in Slavic studies, linguistics, or translation. Recording sessions in which participants did not complete the entire session, either due to a sudden background lighting change or due to issues with holding a pose, were not analyzed. The data obtained from these uncompleted sessions were discarded.

6.2.6 Procedure

Prior to the experiment, the eye-tracker needed to be calibrated. In a first step, participants were asked to adjust the background lighting and to sit comfortably to define a center pose. Calibration was conducted through multiple repetitions of projecting a screen with fixation points (15 to 20) while maintaining a center pose. After the calibration was completed and a face mesh successfully cached, participants began the comprehension test.

The participants' task was to answer the question that immediately followed the stimulus and to press a key indicating the appropriate response. The question was given in the native language of participants. For example Czech native speakers listened to sentences in Bulgarian, Polish, or Russian followed by comprehension checks in Czech. A trial session was conducted prior to starting the experiment. During the trial, the participants were also exposed to stimuli in their own native language to ensure a good understanding of the experimental setup. The fixation dots were presented after each trial, along with a short recalibration procedure, if necessary. Instructions for each phase of the experimental setup were given in the native language of participants. The procedure involving exposure to L1 sentences was used as a control.

The time to first fixation (Duchowski 2017) on the visual field of the direct object has been taken into account in the analysis. The fixations were analyzed in temporal reference to the linguistic events from the auditory stimuli. The x/y coordinates of fixations were compared with the time to the first fixation on a target visual field. Data concerning fixations on the visual target before the verbal component of the audio was played were discarded from the analysis. Such fixations were treated as random and not triggered by information carried by the sentence predicate, which had not yet been perceived.

6.2.7 Data analysis

The descriptive statistics for continuous variables were calculated with the *psych* package in R (R Core Team 2020). Then, a Kruskal-Wallis test was conducted to examine the influence of language on stimulus comprehension with a post-hoc Dunn's test and Holm-Bonferroni correction for pairwise multiple comparisons of the ranked data (Dunn 1964). In the next step, the effects of phonetic distance, surprisal, and language of stimuli on the amount of time until the first fixation on the target visual field were explored by applying a generalized linear model with the *stats* package.

6.3 Results

A Shapiro-Wilk test showed a deviation from the normal distribution. Therefore, non-parametric tests were conducted instead. Basic descriptive statistics for the continuous variables are presented in Table 6.1. In order to conduct more fine-grained analyses and gain insight into pairwise language similarities, an additional grouping variable was added (coded in *xx-yy* format, where *xx* refers to the subjects' native language and *yy* to the language of the stimuli). The basic descriptive statistics for the grouped data are presented in Table 6.4 (section 6.7 Supplementary material).

Table 6.1: Basic descriptive statistics for continuous variables

Variable	*n*	*M*	*SD*	*Mdn*	*Min*	*Max*	*Skew*	*Kurt*	*W*	*p*
Time	2915	0.44	0.41	0.32	0.00	1.63	0.87	−0.30	0.88	< 0.001
Phon. dist.	3851	0.37	0.12	0.38	0.08	0.64	−0.31	−0.51	0.98	< 0.001
Surprisal	3851	6.56	2.22	6.80	0.62	11.20	−0.33	−0.45	0.98	< 0.001

Time - time until the first fixation on the target object, *Phon. dist.* - phonetic distance, *n* - sample size, *M* - mean, *SD* - standard deviation, *Mdn* - median, *Min* - minimum value, *Max* - maximum value, *Skew* - skewness, *Kurt* - kurtosis, *W* - Shapiro-Wilk test, *p* - p-value

The differences in sub-sample sizes vary between time until the first gaze and the independent variables due to analyzing the first gaze on the visual field of the direct object. If an anticipation effect was not discovered, the measurement was excluded from the analysis. In the next step, a contingency table from cross-classifying factors was created (Table 6.2). In addition, a Chi-squared test was conducted. The results showed a significant relation between the subjects' native languages and the languages of stimuli ($\chi^2(9, 3852) = 1286; p < 0.001$) with large effect size (Cramer's V = 0.33). Therefore, an analysis of simultaneous effects of multiple variables was conducted. The significance of the effects of phonetic distance and surprisal and pairs of languages was investigated by a generalized linear model.

Table 6.2: Total distribution of measured trials per L1 and stimulus language with conditional proportions given grand total margin

	Stimuli language				
L1	Bulgarian	Czech	Polish	Russian	Total
Bulgarian	0	428(11.1)	426(11.1)	428(11.1)	1282(33.3)
Czech	257(6.7)	0	257(6.7)	256(6.6)	1027(20.0)
Polish	329(8.5)	320(8.3)	0	325(8.4)	974(25.2)
Russian	278(7.2)	272(7.1)	276(7.2)	0	826(21.5)
Total	864(22.4)	1020(26.5)	959(25.0)	1009(26.1)	100

The nominal variable *language pair* consisted of a set of 12 possible values (four native languages times three languages of exposure) and was changed into a set of dichotomous variables (e.g., bg-pl, bg-cs, bg-ru). The regression analysis was conducted using the step algorithm with the bidirectional search mode (Akaike 1998b). Then the model with the lowest value of the Akaike information criterion was determined (Akaike 1998a). The summary statistics are given in Table 6.3.

Table 6.3: Summary statistics of fitted regression model

L1-stimuli lang.	Estimate	Std. error	t value	Pr(>|t|)
Bulgarian-Czech	0.07	0.03	2.05	0.040
Bulgarian-Russian	0.09	0.04	2.65	0.008
Czech-Bulgarian	0.13	0.04	3.58	< 0.001
Czech-Russian	0.20	0.04	5.01	< 0.001
Polish-Bulgarian	0.08	0.03	2.74	0.006
Polish-Russian	0.09	0.03	2.73	0.006
Russian-Bulgarian	0.14	0.04	4.08	0.001
Russian-Czech	0.07	0.03	2.21	0.027
Intercept	0.45	0.03	16.70	< 0.001
Surprisal	−0.01	0.01	−2.28	0.022

In this analysis, phonetic distance did not appear to be a significant modifier of attention shift and therefore further analyses regard only the information-theoretic notion of surprisal. The effect measured on trials in which subjects listened to Polish sentences did not reach the threshold of statistical significance. The greatest effect was observed in a group of Czech native speakers exposed to Russian stimuli. A strong effect was also measured for Russian native speakers listening to Bulgarian sentences. In contrast, the least significant effect was observed for Bulgarian and Russian native speakers exposed to Czech sentences.

Additionally, a post-hoc Kruskal-Wallis test was conducted to examine differences between surprisal values across the language pairs. The highest values of surprisal were measured for the Bulgarian-Russian cluster, whereas the lowest were found for Czech and Polish, which are both West Slavic languages. Overall, significant differences were found among twelve language pairs ($H = 2654, p < 0.0001, df = 11$) and the effect size was large ($\eta^2 = 0.69$). Therefore, a post-hoc Dunn's test with Holm-Bonferroni correction for pairwise multiple comparisons was conducted. The results showed that 58 out of total 66 pairwise comparisons were significant. For a detailed pairwise analysis see Tables 6.5 and 6.6 (section 6.7).

6.4 Discussion

In line with studies concerning exposure to one's native language only, subjects have shown to be able to identify a predicate from a non-native, but closely related, language. The significance of the effect of surprisal suggests that comprehension of closely related languages can be driven by the notion of stimulus (un)expectedness rather than by a resemblance between phonetic surface forms. The results suggest that subjects were able to identify direct objects better when the Word Adaptation Surprisal between corresponding predicates was low, as opposed to when the phonetic distance between these corresponding word forms was low. This outcome is intriguing, as one might otherwise assume that more similar-sounding words would facilitate intercomprehension to a greater extent than the complexity of sound mappings between languages, which lay listeners may not be aware of. On the other hand, surprisal calculated from regular sound correspondences observed in shared cognates may be able to better distinguish related forms from chance phonetic resemblances, given that unrelated words frequently exhibit chance similarities and that diachronic processes of sound change do not always produce similar-sounding reflexes of inherited phonemes across related languages. Because both cognate and non-cognate verb pairs were included as stimuli, participants needed to first recognize stimuli as cognate to a word in their native language. Therefore, the strong effect of surprisal follows from the fact that it is likely a better predictor of cognacy than a measure of phonetic distance alone.

Interestingly, the asymmetries across the tested languages point to differences in comprehension depending on subjects' L1 and the language of exposure. The intercomprehension pattern reflects the typological division of the Slavic languages only to a limited extent. Strong intelligibility effects were discovered for language pairs that do not belong to the same subgroup of Slavic languages. It was therefore concluded that information received at the verb level was used to direct attention towards the direct object even when exposed to a non-native language.

6.5 Conclusions

In this study, real-time comprehension of spoken stimuli from a non-native yet closely related language was tested in a visual environment. The collected data support the hypothesis that sentence processing in a closely related, non-native language, as reflected by anticipatory eye movements, is driven by the information-theoretic notion of surprisal measured on corresponding predicates. Furthermore, the data obtained in this study exhibit the asymmetrical character of intercomprehension across the four groups of native Slavic speakers.

The data gathered in this study have shown that Slavic native speakers can immediately establish a dependency between the predicate of a sentence and its direct object even if the sentence is perceived in a related, non-native language. The analysis of anticipatory eye movements has shown that individuals listening to sentences in other Slavic languages turn their attention to and begin fixating on the referent objects as soon as they identify a predicate. Furthermore, the relationship between corresponding predicates can be quantified by means of surprisal, which corresponds to the (un)expectedness of stimuli rather than to the degree of resemblance in their surface forms. The study has demonstrated that information extracted at the predicate successfully guides eye movements to an object in a visual setting which satisfies the verb constraints.

This study has shown that comprehension of a predicate causes the attention to shift towards the object before the onset of the referential noun. However, the effect of the language of stimuli moderates this relation depending on the specific stimuli and the subject's L1. Such a finding is in line with previous investigations on the unidirectional character of intercomprehension. Regardless of the discovered asymmetries, this study supports an argument for a surprisal-driven intelligibility effect among speakers of closely related languages.

6.6 Data availability

The experimental data are publicly available in the following (https://osf.io/2wsek/) Open Science Framework repository.

6.7 Supplementary material

Table 6.4: Basic descriptive statistics for continuous variables grouped by L1 and stimuli language: Part 1

Variable	*n*	*M*	*SD*	*Mdn*	*Min*	*Max*	*Skew*	*Kurt*
Bulgarian-Czech (n = 427)								
Time	306	0.42	0.40	0.29	0.00	1.35	0.82	−0.55
Phon. dist.	427	0.40	0.12	0.42	0.18	0.61	−0.51	−0.31
Surprisal	427	7.85	1.15	7.87	5.12	9.42	−1.11	0.79
Bulgarian-Polish (n = 426)								
Time	352	0.41	0.40	0.29	0.00	1.44	0.90	−0.39
Phon. dist.	426	0.40	0.10	0.43	0.23	0.52	−0.60	−1.08
Surprisal	426	4.20	1.01	4.20	2.78	5.56	−0.05	−1.50
Bulgarian-Russian (n = 428)								
Time	322	0.44	0.43	0.30	0.00	1.63	1.01	0.07
Phon. dist.	428	0.42	0.11	0.45	0.29	0.59	0.20	−1.49
Surprisal	428	8.84	1.27	8.83	6.95	11.20	0.21	−0.69
Czech-Bulgarian (n = 257)								
Time	180	0.49	0.42	0.40	0.00	1.51	0.81	−0.32
Phon. dist.	257	0.47	0.06	0.48	0.34	0.55	−0.88	0.16
Surprisal	257	7.88	1.25	7.51	5.96	10.23	0.52	−0.63
Czech-Polish (n = 257)								
Time	201	0.44	0.38	0.31	0.00	1.45	0.77	−0.42
Phon. dist.	257	0.28	0.11	0.24	0.15	0.47	0.19	−1.40
Surprisal	257	2.80	1.77	1.73	0.62	6.03	0.76	−0.90
Czech-Russian (n = 256)								
Time	190	0.54	0.47	0.41	0.00	1.63	0.7	−0.71
Phon. dist.	256	0.34	0.10	0.37	0.17	0.52	−0.14	−0.85
Surprisal	256	8.92	0.81	8.67	8.06	10.72	1.05	0.01

Table 6.5: Basic descriptive statistics for continuous variables grouped by L1 and stimuli language: Part 2

Variable	*n*	*M*	*SD*	*Mdn*	*Min*	*Max*	*Skew*	*Kurt*
Polish-Bulgarian (n = 329)								
Time	242	0.47	0.44	0.34	0.00	1.50	0.82	−0.54
Phon. dist.	329	0.38	0.10	0.40	0.23	0.52	−0.15	−1.51
Surprisal	329	5.01	0.93	5.23	3.49	6.84	0.26	−0.40
Polish-Czech (n = 320)								
Time	235	0.36	0.36	0.22	0.00	1.32	1.05	0.02
Phon. dist.	320	0.29	0.13	0.25	0.09	0.47	−0.21	−1.46
Surprisal	320	6.59	1.66	6.45	3.07	8.39	−0.72	−0.43
Polish-Russian (n = 325)								
Time	223	0.46	0.43	0.33	0.00	1.61	0.86	−0.23
Phon. dist.	325	0.42	0.11	0.45	0.28	0.64	0.56	−0.64
Surprisal	325	6.82	0.92	6.61	5.79	8.80	0.92	−0.38
Russian-Bulgarian (n = 278)								
Time	218	0.50	0.44	0.40	0.00	1.46	0.66	−0.81
Phon. dist.	278	0.42	0.11	0.45	0.29	0.59	0.20	−1.48
Surprisal	278	7.75	1.48	7.59	5.12	9.95	−0.01	−0.95
Russian-Czech (n = 272)								
Time	220	0.44	0.41	0.32	0.00	1.34	0.71	−0.82
Phon. dist.	272	0.34	0.10	0.37	0.17	0.52	−0.14	−0.83
Surprisal	272	7.04	1.29	6.93	5.40	9.51	0.41	−0.95
Russian-Polish (n = 276)								
Time	226	0.39	0.37	0.28	0.00	1.44	0.89	−0.27
Phon. dist.	276	0.25	0.13	0.24	0.08	0.44	−0.02	−1.50
Surprisal	276	4.60	1.00	4.55	2.88	6.42	−0.03	−0.53

Table 6.6: Dunn's test with Holm-Bonferroni correction for pairwise comparisons of surprisal: Part 1

G1	*G2*	*n1*	*n2*	*Est*	*Est 1*	*Est 2*	*Stat*	*p*	*p.adj*
BG-CS	BG-PL	427	426	−1928.43	2609.36	680.93	−25.33	0.000	< 0.001
BG-CS	BG-RU	427	428	488.95	2609.36	3098.31	6.43	0.000	< 0.001
BG-CS	CS-BG	427	257	−48.05	2609.36	2561.32	−0.55	0.584	1.000
BG-CS	CS-PL	427	257	−2195.89	2609.36	413.47	−25.02	0.000	< 0.001
BG-CS	CS-RU	427	256	605.73	2609.36	3215.09	6.86	0.000	< 0.001
BG-CS	PL-BG	427	329	−1608.18	2609.36	1001.18	−19.72	0.000	< 0.001
BG-CS	PL-CS	427	320	−687.93	2609.36	1921.43	−8.37	0.000	< 0.001
BG-CS	PL-RU	427	325	−608.15	2609.36	2001.21	−7.43	0.000	< 0.001
BG-CS	RU-BG	427	278	−111.89	2609.36	2497.47	−1.31	0.192	0.919
BG-CS	RU-CS	427	272	−476.18	2609.36	2133.18	−5.52	0.000	< 0.001
BG-CS	RU-PL	426	276	−1776.29	2609.36	833.07	−20.69	0.000	< 0.001
BG-PL	BG-RU	426	428	2417.38	680.93	3098.31	31.77	0.000	< 0.001
BG-PL	CS-BG	426	257	1880.39	680.93	2561.32	21.41	0.000	< 0.001
BG-PL	CS-PL	426	257	−267.45	680.93	413.47	−3.05	0.002	0.023
BG-PL	CS-RU	426	256	2534.17	680.93	3215.09	28.82	0.000	< 0.001
BG-PL	PL-BG	426	329	320.25	680.93	1001.18	3.92	0.000	0.001
BG-PL	PL-CS	426	320	1240.51	680.93	1921.43	15.08	0.000	< 0.001
BG-PL	RU-BG	426	278	1816.55	680.93	2497.47	21.19	0.000	< 0.001
BG-PL	RU-CS	426	272	1452.26	680.93	2133.18	16.83	0.000	< 0.001
BG-PL	RU-PL	426	276	152.15	680.93	833.07	1.77	0.077	0.536
BG-RU	CS-BG	428	257	−537.00	3098.31	2561.32	−6.12	0.000	< 0.001
BG-RU	CS-PL	428	257	−2684.84	3098.31	413.47	−30.60	0.000	< 0.001
BG-RU	PL-BG	428	329	−2097.13	3098.31	1001.18	−25.73	0.000	< 0.001
BG-RU	PL-CS	428	320	−1176.88	3098.31	1921.43	−14.32	0.000	< 0.001
BG-RU	PL-RU	428	325	−1097.10	3098.31	2001.21	−13.41	0.000	< 0.001
BG-RU	RU-BG	428	278	−600.84	3098.31	2497.47	−7.02	0.000	< 0.001
BG-RU	RU-CS	428	272	−965.13	3098.31	2133.18	−11.19	0.000	< 0.001
BG-RU	RU-PL	428	276	−2265.24	3098.31	833.07	−26.39	0.000	< 0.001
CS-BG	CS-PL	257	257	−2147.84	2561.32	413.47	−21.90	0.000	< 0.001
CS-BG	CS-RU	257	256	653.78	2561.32	3215.09	6.66	0.000	< 0.001
CS-BG	PL-BG	257	329	−1560.14	2561.32	1001.18	16.86	0.000	< 0.001
CS-BG	PL-CS	257	320	−639.88	2561.32	1921.43	6.87	0.000	< 0.001

G1, G2 - groups compared in pairwise test, *n1, n2* - subsamples, *Est* - mean rank difference, *Est 1, Est 2* - mean rank values for the respective group, *Stat* - z-value, *p* - p-value, *p.adj* - adjusted p-value

Table 6.7: Dunn's test with Holm-Bonferroni correction for pairwise comparisons of surprisal: Part 2

G1	*G2*	*n1*	*n2*	*Est*	*Est 1*	*Est 2*	*Stat*	*p*	*p.adj*
CS-BG	PL-RU	257	325	−560.10	2561.32	2001.21	−6.04	0.000	< 0.001
CS-BG	RU-BG	257	278	−63.84	2561.32	2497.47	−0.66	0.507	1.000
CS-BG	RU-CS	257	272	−428.13	2561.32	2133.18	−4.43	0.000	< 0.001
CS-BG	RU-PL	257	276	−1728.24	2561.32	833.07	−17.93	0.000	< 0.001
CS-PL	CS-RU	257	256	2801.62	413.47	3215.09	28.54	0.000	< 0.001
CS-PL	PL-BG	257	329	587.70	413.47	1001.18	6.35	0.000	< 0.001
CS-PL	PL-CS	257	320	1507.96	413.47	1921.43	16.19	0.000	< 0.001
CS-PL	PL-RU	257	325	1587.74	413.47	2001.21	17.11	0.000	< 0.001
CS-PL	RU-BG	257	278	2084.00	413.47	2497.47	21.66	0.000	< 0.001
CS-PL	RU-CS	257	272	1719.71	413.47	2133.18	17.78	0.000	< 0.001
CS-PL	RU-PL	257	276	419.60	413.47	833.07	4.35	0.000	< 0.001
CS-RU	PL-BG	256	329	−2213.92	3215.09	1001.18	−23.89	0.000	< 0.001
CS-RU	PL-CS	256	320	−1293.66	3215.09	1921.43	−13.88	0.000	< 0.001
CS-RU	PL-RU	256	325	−1213.88	3215.09	2001.21	−13.07	0.000	< 0.001
CS-RU	RU-BG	256	278	−717.62	3215.09	2497.47	−7.45	0.000	< 0.001
CS-RU	RU-CS	256	272	−1081.91	3215.09	2133.18	−11.18	0.000	< 0.001
CS-RU	RU-PL	256	276	−2382.02	3215.09	822.07	−24.69	0.000	< 0.001
PL-BG	PL-CS	329	320	920.26	1001.18	1921.43	10.54	0.000	< 0.001
PL-BG	PL-RU	329	325	1000.04	1001.18	2001.21	11.50	0.000	< 0.001
PL-BG	RU-BG	329	278	1496.30	1001.18	2497.47	16.52	0.000	< 0.001
PL-BG	RU-CS	329	272	1132.01	1001.18	2133.18	12.42	0.000	< 0.001
PL-BG	RU-PL	329	276	−168.10	1001.18	833.07	−1.85	0.064	0.512
PL-CS	PL-RU	320	325	79.78	1921.43	2001.21	0.91	0.362	1.000
PL-CS	RU-BG	320	278	576.04	1921.43	2497.47	6.32	0.000	< 0.001
PL-CS	RU-CS	320	272	211.75	1921.43	2133.18	2.31	0.021	0.188
PL-CS	RU-PL	320	276	−1088.36	1921.43	833.07	−11.92	0.000	< 0.001
PL-RU	RU-BG	325	278	496.26	2001.21	2497.47	5.46	0.000	< 0.001
PL-RU	RU-CS	325	272	131.97	2001.21	2133.18	1.44	0.149	0.892
PL-RU	RU-PL	325	276	−1168.14	2001.21	833.07	−12.84	0.000	< 0.001
RU-BG	RU-CS	278	272	−364.29	2497.47	2133.18	−3.84	0.000	0.001
RU-BG	RU-PL	278	276	−1664.40	2497.47	833.07	−17.61	0.000	< 0.001
RU-CS	RU-PL	272	276	−1300.11	2133.18	833.07	−13.69	0.000	< 0.001

G1, G2 - groups compared in pairwise test, *n1, n2* - subsamples, *Est* - mean rank difference, *Est 1, Est 2* - mean rank values for the respective group, *Stat* - z-value, *p* - p-value, *p.adj* - adjusted p-value

Chapter 7

Comprehension of Idiomatic Phrases

Continuing the extension of the analyzed units, this chapter focuses on comprehension of idiomatic phrases. It involves a series of complex translation tasks in which the stimuli included multiword expressions in Polish and Russian.

The results of this study were presented during the Europhras 2021 conference at UCLouvain, Louvain-la-Neuve, Belgium, on 6-9.09.2021 and the manuscript is currently under review.

7.1 Introduction

The comprehension of idiomatic phrases in a closely related language is a challenging task, especially if a perceiver has no previous training in philology, translation studies, or linguistics. Additional challenges are encountered in the auditory modality. However, the investment of this extra cognitive effort may be rewarded when an idiomatic meaning is successfully understood and a humorous or surprising cross-cultural meaning is discovered. The ability to understand the figurative meaning of an expression in a foreign language often corresponds with high linguistic proficiency and so far has been mainly investigated for second language learning (Boers and Demecheleer 2001; Cieślicka et al. 2017; Kovecses and Szabco 1996). Thus, the comprehension of idiomatic phrases from closely related yet unfamiliar languages is far more complicated. This phenomenon involves phonological, lexical, and syntactic correspondences between similar languages and not only raises questions of communicative competence, but also touches upon the abilities and limitations of linguistic transfer based on various associations. A possible strategy is to conceptualize the meaning of each constituent of an idiomatic phrase in Lx and, upon encountering difficulty, try to find a comprehensible lexeme in it (e.g., a cognate component) and then collocate the rest on the basis of native language (L1) structure. Another strategy involves the surface representation of a target phrase and attempts to match phonologically close units from one's L1 and the foreign language Lx. But even in one's native language, the comprehension of idioms requires a certain level of linguistic and cultural competence. Therefore, it is worth investigating what kind of cross-lingual references play a key role in equivalent matching and how to measure the cues which are informative in such transition processes.

7.1.1 Aims and hypotheses

To test intercomprehension of idiomatic phrases, a translation experiment involving native speakers of Polish and Russian without any formal training in linguistics, philology, or Slavistics was conducted. This study combines cross-lingual phrase (un)expectedness with measurable phonetic, lexical, and syntactic distances between idiomatic expressions in order to answer the following research questions.

Firstly, it is assumed that if the idiomatic meaning of a target phrase is not transparent enough, subjects tend to understand it literally. The phenomenon driving such a preference is the phonological resemblance between an idiomatic target and its literal translation. Hence, it is hypothesized that the lower the mean phrase word adaptation surprisal (henceforth: WAS) between the stimulus phrase and the literal translation, the stronger subjects' preference is to select the literal interpretation.

Secondly, it is assumed that cognates can serve as cues when matching translations, which makes participants preferentially select expressions with recognizable lexical material. Hence, to quantify the influence of cognate-based correspondence, the measure of normalized lexical distance between the stimulus phrase and the equivalent was introduced. It is hypothesized that lower normalized lexical distances lead to stronger preferences for translations that share a cognate lemma with the stimulus, but still diverge from the idiomatic meaning on a phrase level.

Thirdly, the question of linear syntactic correspondences between idiomatic phrases was addressed by referring to the compositional vs. non-compositional dichotomy. It is assumed that a syntactic distance between the stimulus and the respective translation influences participants' preference. It is hypothesized that the lower the syntactic distance, computed as the normalized sum of insertion and deletion operations, the stronger the preference for the semantic equivalent of the stimulus phrase in both translation directions.

7.1.2 Related work

In intercomprehension, speakers of closely related languages opt to use their mother tongue under the assumption that similarities between their languages will allow for relatively unhindered communication. Such a preference to use one's L1 might be motivated by the proximity of the languages in question, or the lack of a suitable lingua franca. Previous studies on the mutual intelligibility of closely-related languages have shown that the strategy of using one's own L1 can be successful in practice (Bezooijen and Gooskens 2007; Stenger et al. 2017; Gooskens et al. 2018). As pointed out by (Boers and Demecheleer 2001), comprehension problems caused by cross-cultural differences will mostly be confined to situations where distant cultures meet. However, members of language communities sharing both similar languages and cultures, who would ordinarily exhibit high degrees of mutual intelligibility, may nevertheless encounter challenges when exposed to various multi-word expressions. The existence of syntactic nuances, false friends, and contextual constraints complicate fluent communication in multilingual situations.

The question of idiom familiarity in a setup similar to this study was raised by (Schweigert 1986), who measured the reading times of sentences with varying degrees of idiomaticity. A different approach was taken by (Cacciari and Tabossi 1988) in their study on comprehension of idioms in cross-modal priming and contextually constrained predictability. Most of the newer studies in the field of phraseology are conducted on written material rather than spoken data (Rubio 2020). Indeed, the comprehension of phraseological expressions in the auditory modality presents an additional complexity. Furthermore, previous investigations of idiom comprehension and translation mainly involved reaction times or gaze tracking (Kessler et al. 2020), rarely taking an information-theoretic approach (Petrova and Mikhailovskaya 2014).

Debates on how best to render idiomatic phrases in translation are vibrant among professional translators. A frequently cited difficulty is to strike a compromise between the cultural and non-linguistic information conveyed by an idiom and the selection of an accurate equivalent in the target language. Hence, translation of idioms, even between closely-related languages, poses a challenge for both human and machine translators (Hejwowski 2004; Berman 2000; Jadlovsk 2020). The notion of quantifying idiomatic phrases in a cross-lingual perspective was already raised by (Cieślicka 2015).

Several strategies relevant to comprehension of idioms have been proposed. Some of these include a cognitive effort component but neglect the notion of phrase (un)expectedness. In on-line idiom comprehension situations, as pointed out by (Vega-Moreno 2001), the computation of word meanings is exceptionally costly and, according to the Communicative Principle of Relevance (Grice 1975), should be avoided if an easier solution is available.

Therefore, an additional cross-lingual component was considered by the application of the information-theoretic measure of surprisal, which correlates to cognitive processing effort (Hale 2001; Levy 2008). In the proposed experimental setup, the higher the measured surprisal is, the more effort is required for comprehending the target phrase, presupposing a certain degree of Slavic mutual intelligibility (Golubović and Gooskens 2015; Golubović 2016; Jágrová et al. 2019).

7.1.3 Approaches to comprehension of idioms

The classical non-compositional model describes idiomatic phrases as strings in the lexicon whose meaning is not derived from the sum of their component words. This view, however, cannot account for alternations which preserve the non-figurative meaning despite syntactic modifications. Word order constraints in such expressions stem from their historical development, hierarchical structure, and semantics (Cutler 1982; B. Fraser 1970; Nunberg et al. 1994; Cronk and Schweigert 1992). While mono-lingually syntactic alternations can still preserve an idiomatic meaning with respect to a canonical form, in a multilingual setup the comprehension may be affected by a range of syntactic modifications. In a translation scenario, the understanding of an isolated idiomatic phrase can depend on its syntactic structure and on the degree to which this matches its counterpart in another language. The proposed quantification of syntactic alternations from a cross-lingual perspective may be advantageous for interpreting the results of isolated idiom translation between mutually intelligible languages. In this work, the normalized InDel (Heeringa et al. 2018) distance was calculated. It refers to a syntactic distance measure, which corresponds to the degree of internal syntactic change calculated as the normalized sum of insertion and deletion operations between corresponding phrases.

Viewing idioms as independent lexical items (Bobrow and S. M. Bell 1973) has led to the the Literal First Hypothesis, according to which the literal meaning of idioms precedes the idiomatic or figurative meaning of a phrase. This concept has been questioned by several psycholinguistic interpretations (Gibbs 1980). In contrast to the theory introduced in (Bobrow and S. M. Bell 1973), the Simultaneous Processing Hypothesis proposed by (Swinney and Cutler 1979) has been supported by experimental results which suggest that the processing time of literal strings is not longer than that of idiomatic ones (Vega-Moreno 2001). This laid the foundation for the Parallel Processing Hypothesis, according to which both processes (literal and idiomatic meaning retrieval) run simultaneously or, as in its more radical version, the meaning of an idiomatic phrase does not need to be activated at all. Non-compositional accounts view idioms differently and argue instead that the idiomatic meaning of phrases is not entirely arbitrary. The investigation into non-arbitrariness of idiomatic phrases often requires taking a diachronic perspective and entails etymological research on phrasal components. In this line, (Nunberg et al. 1994) proposed that idiomaticity is conveyed on the level of semantics rather than syntax, and called this approach Decompositional Hypothesis. This view considers the role of specific words composing an idiomatic phrase. In response to this theory, the proposed experimental setup involves gradual changes on the syntactic level of the stimuli and measures their understandability to lay translators by implementing distances measures between Polish and Russian phrase equivalents.

Alternatively, Phrase-induced Polysemy does not reject the concept of string-like idioms stored in the mental lexicon and proposes a certain threshold at which the understanding of a phrase changes from a literal to an idiomatic one. Such a threshold is often called the idiomatic key (Vega-Moreno 2001). Before this point is reached, both interpretations occur simultaneously. Given that they originate from common roots in an ancestral language and still share a semantic field in both daughter languages, cognates and partial cognates are the easiest units to understand in the cross-lingual perspective. In this experiment, the idiomatic key is defined in terms of a lexical distance measure which quantifies the proportion of cognates and partial cognates in a phrase pair.

7.2 Methods

To quantify cross-lingual phraseological differences idiomatic phrases were explored in isolation, e.g., occurring as entries in a bilingual dictionary. The subjects are native speakers untrained in translation. Subjects' preferences are revealed in both open and closed tasks. In the open task, the participants were asked to write their own translation of 43 idiomatic phrases that were presented to them in the auditory modality, i.e., Polish native speakers listened to Russian phrases and were asked to translate them into Polish, and vice versa. In the closed task, participants were given three possible equivalents of the target phrases in their native language: (1) a lemma-based equivalent that shares a lexical (cognate) component still differing in the rest of the phrase; (2) a literal translation of the source that diverges from the target idiomatic interpretation; and (3) a proper semantic equivalent from a phraseological dictionary entry (Fedorov 1995; Lukszyn 1998; Chlebda 2016). All recorded phrases had corresponding semantic equivalents in the Polish Russian phraseological dictionary; phrases without semantic equivalents in both languages were not considered.

7.2.1 Audio stimuli

In total, 43 target idiomatic phrases were tested. The audio samples were read by female native speakers of the respective language (Russian or Polish) and recorded in an acoustically controlled environment at a 48 kHz sampling rate to uncompressed format.

7.2.2 Participants

In total, 100 participants (50 native speakers per language) took part in the translation study. The subjects reported no hearing disorders, and no formal education in linguistics, Slavistics, or translation studies. They were recruited via an online crowd-sourcing platform and were paid for their participation.

7.2.3 Normalized InDel

When processing a sentence or a phrase in a closely related language, listeners or readers may expect some words to be added or deleted in order to yield the closest possible translation into their native language. The normalized InDel is a measure of word insertions and deletions (Heeringa et al. 2018) to quantify linear syntactic alternations. Larger numbers of added or deleted words imply more negative effects on comprehension of a phrase in a related language. As an illustration of InDel computation, Polish and Russian semantic equivalents with idiomatic meaning 'intoxicated, drunk' (e.g., English 'three sheets to the wind') are presented in Table 7.1.

Table 7.1: Example of the alignment for InDel distance computation between two semantic equivalents

Language	1	2	3	4	5	Literal translation
Polish	pijany	jak	bela			drunk as a log
Russian	пьяный			в	дым	drunk into smoke

For InDel calculation, the phrases are aligned, so the corresponding syntactic elements occur in the same column, while gaps represent elements missing in one of the languages. For instance, the Polish *pijany* is aligned with the Russian *пьяный*, both meaning 'drunk', whereas the remaining words in each language do not have equivalents in the other. Thus, two deletions can be observed (Polish *jak* 'as' and *bela* 'a log') and two insertions (Russian *в* (v)'into' and *дым* (dym) 'smoke'). This alignment results in an InDel value of 4. Normalizing the InDel distance by the number of the alignment slots (in this case 5), a normalized InDel (henceforth: nInDel) equals to 0.8 (Gooskens and F. Swarte 2017) - that is, 80% of the alignment consists of elements which have no equivalent in the corresponding phrase. Table 7.2 presents the mean nInDel distance values (from source to target alignments) for each condition in both directions.

7.2.4 Mean phrase word adaptation surprisal

The information-theoretic notion of surprisal (Shannon 1948) was applied to model the predictability of a particular cross-lingual correspondence for a given language pair. The basic assumption is that a native

Table 7.2: Mean nInDel distance between Polish and Russian for each condition

Stimulus	Subjects' L1					
	Polish			Russian		
	LEM	SEM	LIT	LEM	SEM	LIT
Polish				0.44	0.56	0.18
Russian	0.49	0.56	0.20			

LEM - lemma-based equivalent, *SEM* - semantic equivalent, *LIT* - literal translation

speaker of L1 understands a word from a related Lx and so can predict a word of L1 which is the best equivalent for the word of Lx. In the case of cognates, such a prediction can be based on regular sound correspondences between the related lexemes. The more cognate pairs are shared by two languages, the better intercomprehension can be expected. However, phonetic and morphophonological aspects of cognate words are subject to diachronic changes and may no longer be transparent to a language user. To counterbalance the cognate recognition effect, the tested phrases contained cognates as well as non-cognates. During the experiment, the subjects were exposed to the entire set of cognates and non-cognates constituting the idiomatic phrases, hence this measurement method was based on all aligned word pairs, including referents both with and without regular sound correspondences (Moberg et al. 2007). Mean phrase word adaptation surprisal (WAS) corresponds to the sum of the phone adaptation surprisal values and is calculated by the following equation:

$$\mathrm{WAS}(L1 = c1|L2 = c2) = -\log_2 P(L1 = c1|L2 = c2)$$

where *L1* = native language, *c1* = phone of the native language, *L2* = stimulus language, and *c2* = phone of the stimulus language. Individual phone transformation probabilities were extracted with Lidstone smoothing from the corpus of phonetically aligned cognate words used in the study, yielding probabilities of encountering individual phones given their aligned equivalent in the other language (e.g., the probability of /ʨ/ in Polish given /t^j/ in Russian). WAS is computed in bits according to these phone transformation probabilities (Table 7.3), and normalized by the number of alignment slots for the particular word pair (Mosbach et al. 2019).

Table 7.3: Example of calculation of phoneme-based mean phrase WAS in bits

RU orthography	отплатить	той		же	монетой
PL orthography	odpłacić	tą	samą		monetą
RU IPA	/ətpłɐtʲˈitʲ/	/toɪ/		/zɨ̞/	/mɐnʲˈetəɪ/
PL IPA	/ɔtpwˈatɕitɕ/	/tɔw̃/	/sˈamɔw̃/		/mɔnˈɛtɔw̃/
WAS	0.84	1.48	5.49	8.20	1.41
Phrase surprisal			3.48		

It is assumed that higher normalized WAS values cause more intercomprehension difficulties. Mean phrase WAS values were calculated by averaging the nWAS values for each alignment position, including both aligned word pairs and words aligned with gaps. The nWAS values for words aligned with gaps were calculated using the surprisal of each segment in the word aligned with a gap character. Such a procedure provides a quantification of the overall (un)expectedness of the respective phrase. Table 7.4 presents the mean phrase WAS (in bits) between Polish and Russian for the three conditions.

Table 7.4: Mean phrase WAS in bits between Polish and Russian for each condition

Stimulus	Subjects' L1					
	Polish			Russian		
	LEM	SEM	LIT	LEM	SEM	LIT
Polish				5.82	6.12	4.74
Russian	5.78	6.01	4.08			

LEM - lemma-based equivalent, *SEM* - semantic equivalent, *LIT* - literal translation

7.2.5 Lexical distance

It was expected that a large proportion of cognates in the aligned phrases facilitate comprehension of the stimulus. As non-cognates (etymologically unrelated words) tend to be unintelligible to listeners without prior knowledge of the stimulus language, it was assumed that larger proportions of non-cognates would impede intercomprehension. According to (Gooskens 2019), the percentage of non-cognates determines the lexical distance between closely-related languages. Phonologically close but semantically distant equivalents, so-called false friends, e.g., Russian *урод* [ʊrˈot] ‘monster’ and Polish *uroda* [urˈɔda] ‘beauty’, should be treated with special attention, since they may cause even larger difficulties than non-cognates.

Aligned phrase pairs are scored for lexical distance by assigning distances to word-form pairs of each type. Non-cognates and false friends are assigned a distance of 1, and cognates have 0 distance. Words in the stimulus without aligned equivalents were also assigned a distance of 1. The lexical distance for a phrase is then yielded by dividing the sum of word pair distances by the number of words contained in the phrase.

The lexical distances can be asymmetric due to synonyms which may be cognate with the target word. For example, Russian *собака* [sɐbˈakə] ‘dog’ translates to Polish *pies* [pʲjɛs] ‘dog’, forming a non-cognate pair. Nevertheless, a Russian listener understands the spoken Polish word *pies* because a phonetically close synonym *пёс* [pʲos] ‘dog’ exists in Russian. As the Russian word *собака* does not have any cognate synonym in Polish, a Polish listener cannot understand it without prior knowledge of Russian.

With an increasing percentage of non-cognates, partial cognates, and false friends, the subjects may encounter more difficulty in understanding a phrase in a related but unfamiliar language. Table 7.5 shows an example calculation of lexical distance. Table 7.6 presents the mean lexical distances between Polish and Russian across the three conditions as the aggregate distance for all parallel phrases in the corpus for each condition in both directions.

Table 7.5: Example of the alignment for lexical distance calculation

Polish	koń	w	mydle
Russian	лошадь	в	мыле
English	horse	in	soap
Cognate/non-cognate	non-cognate	cognate	cognate
Scores	1	0	0
Lexical distance	0.33		

Table 7.6: Mean lexical distance between Polish and Russian across conditions

Stimulus	Subjects' L1					
	Polish			Russian		
	LEM	SEM	LIT	LEM	SEM	LIT
Polish				0.62	0.69	0.52
Russian	0.62	0.68	0.42			

LEM - lemma-based equivalent, *SEM* - semantic equivalent, *LIT* - literal translation

7.2.6 Experimental procedure

Before the translation task, participants were exposed to a trial session to familiarize themselves with the experimental setup. The trial session included examples of both closed and open set tasks. During the trial session the participants could adjust the volume to a comfortable level. They were asked to wear headphones for better sound quality. The experimental sessions consisted of two parts. The open translation task preceded the closed set question task to avoid a bias towards already seen equivalents. The participants were instructed to provide a translation of each idiomatic phrase into their L1. In the first part, the stimuli were presented only in the auditory modality and subjects were instructed to provide a written translation. The listening task was then followed by the closed set question, in which participants were given on the screen three possible written equivalents in their native language. The audio recording of the phrase was automatically played at the beginning of each question, and subjects could replay the audio if they desired. The participants had to provide an answer before moving on to the next question. They were not asked to rate their confidence in their responses. The entire session lasted around 20 minutes. To avoid fatigue, subjects were allowed to self-pace themselves through the experiment. However, they were instructed to remain at the experimental screen and not to use external resources for the translation task. A full-screen tracking plugin was active during the session. The experiment had to be completed in one uninterrupted session. After the completion of the entire experiment, the subjects were paid for their participation via a crowd-sourcing platform. They were not provided with feedback on their performance.

7.3 Results

In order to answer the research questions, multinomial logistic regression models were run with the significance level set at $\alpha = 0.05$. The results from the open set tasks were classified into four main categories: (1) translations motivated by phonetic equivalence, where a strong sound resemblance between the stimulus phrase and the provided translation was taken as a cue; (2) identification of a specific component of the stimulus phrase matching with its L1 equivalent, where the answers were motivated by the correct identification of one unit of the stimulus phrase and then complemented with plausible collocates, sometimes resulting in an idiomatic phrase in the subjects' L1; (3) dictionary-based equivalent, often close to a literal translation and not directly referring to the idiomatic interpretation of the stimulus phrase; and (4) other type of equivalence, neither clearly motivated by phonetic correspondence nor by semantic resemblance.

The results of the closed set translation section were quantitatively analyzed with respect to the phonetic surprisal and the syntactic and lexical distances between the stimuli and the selected phrases. Such an approach helps to understand whether native speakers' preferences for selecting a particular equivalent are guided by semantics, lexical similarity, or a surface-oriented interpretation.

7.3.1 Open set

In total, in the Russian to Polish direction, 2528 translations were classified and 2451 in the Polish to Russian direction. Translations that were incomplete, typed as a random string of symbols, question marks, or ellipses were discarded from the set of analyzed data.

7.3.2 Polish to Russian

The similarity of syntactic structure and straightforward identification of cognates allowed for a consistent recognition of the idiomatic phrase pair *odpłacić tą samą monetą* (Polish) and *отплатить той же монетой* (Russian), whereby 72% of the answers belonged to the semantically equivalent category. The provided equivalents preserved the idiomatic meaning of the target phrase and maintained its syntactic structure. Such an outcome, however, was relatively rare in the remaining phrases, which were syntactically more divergent. Overall the interpretation based on phonetic correspondences between target phrase and selected translation accounted for 31% of all responses. The second most frequent strategy was cognate identification combined with strong collocates in Russian. This type of responses accounted for 23% of all equivalents in the free translation task. Only 10% of translations preserved the idiomatic meaning of the target phrase. The remaining 36% of responses were unclassified.

7.3.3 Russian to Polish

Only 15% of open set task responses were classified as dictionary-based equivalents with a preserved idiomatic meaning. The most frequent pattern in the free translation part was inspired by the surface phonetic representation of the target phrase (26%). Another common technique was lemma-based identification accompanied by frequent collocates in Polish, which constituted 24% of all answers. The remaining 35% of translations given in the open set task could not be classified. A closer look at the translation pairs revealed that the exception to these patterns were phrase pairs with equal syntactic structure in both languages supplemented by cognate tokens, e.g., Russian *здоров как бык* and Polish *zdrowy jak byk*, where 85% of open set translations belonged to the third category which preserved the idiomatic meaning of the stimulus phrase. The difficulty of such a task motivates the technique for providing equivalents based on phonetic correspondences and cognate recognition rather than the dictionary entries.

7.3.4 Closed set

The results from the closed set were analyzed by modeling the impact of lexical indices on the probability of choosing one of three given translation equivalents: (1) lemma-based translation, where the equivalent and the target phrase share a lemma; (2) literary translation, based on a direct translation of the phrase constituents but lacking the idiomatic dimension; and (3) semantic-based equivalent taken from the phraseological dictionary, where often the components of the stimulus phrase and the translation were different, but preserved the idiomatic meaning. Table 7.7 shows a summary of the selections in each direction.

The overall patterns of translation preferences in the closed set task do not match the tendencies from the open set task. Being provided with three options, more participants tended to select one of the non-literal translations of the target phrase, as compared to the open set task in

Table 7.7: Participants' choices in the closed set task

Direction	LEM	LIT	SEM
Russian to Polish	21%	51%	28%
Polish to Russian	27%	40%	33%

LEM - lemma-based equivalent, *SEM* - semantic equivalent, *LIT* - literal translation

which relatively few non-literal responses were recorded. However, the literal translation equivalent was still the single most dominant selection, in both directions. The literal equivalents without idiomatic meaning accounted for more than half of all selections in the Russian to Polish direction, and 40% in the opposite direction. Since the translations in both language groups diverge, the detailed analysis of the results should be conducted for both directions separately.

7.3.5 Polish to Russian

In the Polish to Russian direction, the likelihood of choosing the LEM, LIT, and SEM equivalents was predicted by mean phrase WAS, nInDel and lexical distance. For the comparison, first the LEM was used as the reference category coded (0) and another model was run with LIT as reference category coded (0) to allow a contrastive analysis between LIT and SEM. Compared to the null model (with no applied predictors), the tested model showed significant improvement based on likelihood ratio test ($\chi^2(6) = 515.89, p < 0.001$), however Pearson ($\chi^2(224) = 4238.86, p < 0.001$), and deviance ($\chi^2(224) = 4127.79, p < 0.001$) tests both indicated that the model does not fit the data well.

All predictors included in the model were significant: mean phrase WAS ($\chi^2(2) = 65.33, p < 0.001$), nInDel ($\chi^2(2) = 58.65, p < 0.001$), and lexical distance ($\chi^2(2) = 28.60, p < 0.001$). The odds of choosing LIT compared to preference of LEM equivalent decreased, with increase in mean phrase WAS ($B = -0.572, SE = 0.072, p < 0.001, OR = 0.564$) and nInDel, ($B = -0.681, SE = 0.277, p = 0.014, OR = 0.506$), but increased with increase in lexical distance ($B = 1.63, SE = 0.331, p < 0.001, OR = 5.092$). The odds of choosing SEM compared to choosing LEM decreased, with increase in mean phrase WAS ($B = -0.264, SE = 0.067, p < 0.001, OR = 0.768$), but increased with increase in nInDel, ($B = 1.267, SE = 0.261, p < 0.001, OR = 3.552$), and lexical distance,

Table 7.8: Model predicting the type of chosen equivalent based on mean phrase WAS, nInDel, and lexical distance in Polish to Russian translation

Comparison	Predictor	*B*	*SE*	*Wald*	*df*	*p*	*Exp(B)*
	Intercept	2.239	0.177	160.154	1	< 0.001	
LEM (0)	Mean WAS	-0.572	0.072	62.432	1	< 0.001	0.564
LIT (1)	nInDel	-0.681	0.277	6.029	1	0.014	0.506
	Lex. dist.	1.628	0.331	24.211	1	< 0.001	5.092
	Intercept	0.332	0.181	3.375	1	0.066	
LEM (0)	Mean WAS	-0.264	0.067	15.351	1	< 0.001	0.768
SEM (1)	nInDel	1.267	0.261	23.580	1	< 0.001	3.552
	Lex. dist.	1.261	0.307	16.915	1	< 0.001	3.528
	Intercept	-1.907	0.168	129.212	1	< 0.001	
LIT (0)	Mean WAS	0.308	0.068	19.836	1	< 0.001	1.361
SEM (1)	nInDel	1.948	0.263	54.786	1	< 0.001	7.017
	Lex. dist.	-0.367	0.322	1.297	1	0.255	0.693

($B = 1.261, SE = 0.307, p < 0.001, OR = 3.528$). The odds of choosing SEM compared to choosing LIT increased with increase in mean phrase WAS ($B = 0.308, SE = 0.069, p < 0.001, OR = 1.361$), and nInDel, ($B = 1.948, SE = 0.263, p < 0.001, OR = 7.017$). The lexical distance did not differentiate between SEM and LIT ($B = -0.367, SE = 0.322, p = 0.255$). The details are presented in Table 7.8. Overall, choosing the LEM equivalent was correctly predicted by the model only in 11% of cases. Preference towards the LIT equivalent was correctly predicted 71.3% of the time; whereas choosing SEM equivalent was correctly predicted 59.3% of the time. The mean correct prediction was 50.5%.

7.3.6 Russian to Polish

A similar analysis was conducted to interpret the results in the Russian to Polish direction. Compared to the null model, the tested model showed significant improvement based on likelihood ratio test ($\chi^2(6) = 773.85, p < 0.001$), however Pearson ($\chi^2(218) = 4337.91, p < 0.001$), and deviance ($\chi^2(218) = 3933.85, p < 0.001$) tests indicated that the model does not fit the data well. All predictors included in the model were significant: mean phrase WAS ($\chi^2(2) = 27.87, p < 0.001$), nInDel ($\chi^2(2) = 114.18, p < 0.001$) and lexical distance ($\chi^2(2) = 61.63, p < 0.001$).

Table 7.9: Model predicting the type of chosen equivalent based on mean phrase WAS, nInDel, and lexical distance in Russian to Polish translation

Comparison	Predictor	*B*	*SE*	*Wald*	*df*	*p*	*Exp(B)*
	Intercept	3.580	0.219	267.458	1	< 0.001	
LEM (0)	Mean WAS	-0.287	0.061	22.470	1	< 0.001	0.750
LIT (1)	nInDel	-1.276	0.243	27.469	1	< 0.001	0.279
	Lex. dist.	-1.900	0.253	56.571	1	< 0.001	0.150
	Intercept	0.972	0.245	15.749	1	< 0.001	
LEM (0)	Mean WAS	-0.059	0.063	0.872	1	< 0.350	0.943
SEM (1)	nInDel	1.243	0.240	26.749	1	< 0.001	3.466
	Lex. dist.	-1.502	0.263	32.532	1	< 0.001	0.223
	Intercept	-2.609	0.185	197.892	1	< 0.001	
LIT (0)	Mean WAS	0.229	0.057	15.970	1	< 0.001	1.489
SEM (1)	nInDel	2.519	0.243	107.284	1	< 0.001	1.489
	Lex. dist.	0.398	0.236	2.850	1	0.091	1.257

The odds of choosing LIT compared to choosing LEM decreased, with increase in mean phrase WAS ($B = -0.287, SE = 0.061, p < 0.001, OR = 0.750$), nInDel ($B = -1.276, SE = 0.243, p < 0.001, OR = 0.279$), and lexical distance ($B = -1.900, SE = 0.253, p < 0.001, OR = 0.150$). The odds of choosing SEM compared to choosing LEM decreased, with increase in lexical distance ($B = -1.502, SE = 0.263, p < 0.001, OR = 0.223$), but increased with increase in nInDel ($B = 1.243, SE = 0.240, p < 0.001, OR = 3.466$). The mean phrase WAS did not differentiate between SEM and LEM ($B = -0.059, SE = 0.063, p = 0.350$). The odds of choosing SEM compared to choosing LIT increased with increase in mean phrase WAS ($B = 0.229, SE = 0.057, p < 0.001, OR = 1.257$) and nInDel ($B = 2.519, SE = 0.243, p < 0.001, OR = 12.412$). The lexical distance did not differentiate between SEM and LIT ($B = 0.398, SE = 0.236, p = 0.091$). Details were provided in Table 7.9.

7.4 Discussion

The data obtained from the translation experiments showed different strategies in providing translation equivalents across the open set and closed set tasks. In the open set, most of the answers were motivated by a phonetic interpretation of the stimulus phrase, with the notable exception of obvious semantic equivalents which shared both cognate cues and exhibited similar syntactic structure. Another frequent strategy involved translation guided by cognate identification and strong collocates in the subjects' L1, which often resulted in a translation that diverged from the idiomatic meaning of the stimulus phrase. This tendency can also be explained by phonetic and phonological neighborhood density which can supplement the process of equivalent matching in translation tasks. The results of the closed task contribute to understanding the preference of native speakers to select a particular equivalent guided by semantics, lexical similarity, or surface oriented interpretation. The outcomes from the open task, by categorizing the translations into several groups, can provide insight into phrase intelligibility. Several dependencies were discovered across the two groups of tested subjects. In both directions of translation, the proposed measures appeared to be significant predictors of lay translators' preferences. However, across the L1 groups, the research hypotheses were not supported to equal extent. Due to these differences across the subject groups, the interpretation of the results is conducted separately for both directions of translation.

7.4.1 Polish to Russian

The first hypothesis referred to a preference to select a literal translation equivalent when it exhibits a lower mean phrase WAS. The obtained data support the hypothesis that the mean phrase WAS is an accurate predictor of subjects' preference towards selecting the literal translation equivalent without idiomatic meaning. The second hypothesis predicted that lower lexical distance would correspond to a higher preference for the equivalent sharing a cognate lemma. A strong effect was observed, suggesting that a decrease in lexical distance indeed correlates with an increased preference for selection of the equivalent which shares a cognate word but does not entirely correspond to the phrase idiomaticity. The third hypothesis attempted to explain the preference for the semantic equivalent through the syntactic distance between the two phrases computed as nInDel measure. The data do not support this hypothesis. In fact, the relation between the preference for the semantic equivalent and the nInDel measure appears to oppose this assumption. That is, the higher nInDel values, the stronger the preference for the phrase with actual idiomatic semantics.

7.4.2 Russian to Polish

The first hypothesis was also supported in the Russian to Polish direction. The mean phrase WAS was an accurate predictor of the preference for the literal translation equivalent. The comparison with the LEM and SEM variants, however, did not reach the threshold of statistical significance. The second hypothesis referring to cognate-based selection predicted by lexical distance was rejected in the Russian to Polish direction, namely, the opposite effect to that in the Polish to Russian direction was observed. Comparison with the LIT and SEM conditions did not reach the threshold of statistical significance. The third hypothesis was also rejected, similar to the Polish to Russian direction. The greater the syntactic divergence between the target phrases, the stronger the preference for the selection of the idiomatic equivalent.

7.5 Conclusions

In this study, an attempt to quantify intercomprehension of idiomatic phrases in closely related languages was made with the application of information-theoretic measures. The results contribute to understanding of the importance of phonetic, lexical, and syntactic cues in a process of equivalent selection for isolated idiomatic phrases. The conclusions are drawn on the outcomes from a quantitative analysis of the closed set results. An interesting asymmetry was observed in the translation direction. Different strategies were discovered in the open set and in the closed set tasks. The effect of strong surface phonetic similarities of phrases seems to motivate the equivalent matching, especially with respect to literal translation equivalents. Cognate lemma-based identification also seems to play a role in naïve translation, with lexical distance often serving as an idiomatic key. Overall, the data suggest that phonetic, lexical, and syntactic measures of idiomatic phrase pairs can provide an explanation for strategies used by native speakers of closely related languages in the selection of phrasal correspondences.

Even though comprehension of idiomatic phrases is known to be difficult, the gathered data show how idiomatic expressions tend to be comprehensible for listeners whose L1 is closely related to the stimulus language. The preference to choose a particular type of translation equivalent, which serves as an indicator of intercomprehension, can be predicted by phonetic, lexical, and syntactic similarities. The nature of mutual intelligibility is asymmetric, therefore different idiomatic keys were discovered for Polish and Russian native speakers. Directions for future work might involve an experimental design with more language pairs. Examining less closely related languages would entail a paradigm shift from intercomprehension to L2 studies and could also create an interesting parallel to this experiment. The analysis would also benefit from the implementation of additional predictors as well as further data collection.

7.6 Data availability

The experimental material along with the datasets obtained in the open set and closed set translation tasks are publicly available in the following (https://osf.io/abrdc/) Open Science Framework repository.

Chapter 8

General Discussion

In this work, a series of five studies aimed to investigate the spoken intercomprehension among speakers of closely related languages. In total, 629 subjects took a part in various tests primarily in the auditory modality. The participants were native speakers of four Slavic languages: Bulgarian, Czech, Polish, and Russian. Such a selection allowed for investigating receptive multilingualism across the typological and genealogical groupings of the Slavic family and finding correspondences, as well as asymmetries, in intercomprehension. The outcomes of the studies have shed light on auditory processing of closely related language. Furthermore, the application of several experiments addressing various units of linguistic analysis revealed the task dependencies in cross-lingual comprehension of speech.

8.1 Structuralist perspective

In a prism of studies on receptive multilingualism, foreign language comprehension varies depending on decoded information as well as on investigated subsystem of a language. Starting from the segmental phonetic level, the proposed LADO study focused on the spectral and temporal properties of segments shared by the phonological inventories of four Slavic languages. It aimed to discover which cues (language-specific stress distribution patterns, vowel space overlap, pseudoword length, or logatome identification surprisal) are relevant to lay listeners for determining a speaker's language of origin and to estimate an importance of fluency in non-Slavic languages on performance in a quasi-forensic setup. On suprasegmental level, the hypothesis regarding stress differences caused by disyllabic and trisyllabic sequences exhibiting various stress patterns has not been confirmed. The gathered data suggest that stress distribution is not a discriminable factor. Therefore, a position of lexical stress is not informative enough to discriminate a linguistic origin of speakers coming from languages with free and fixed stress placement.

Lexical processing of spoken input coming from a closely related language has been investigated in a short-term priming paradigm. A cognate facilitation effect has been discovered in a multilingual lexical decision task. The ability of Slavic native speakers to immediately recognize the associative link between a non-native Slavic prime and a target in their Slavic L1 contributes to a strong intercomprehension effect. Intelligibility of the four Slavic languages on the sentential level has been investigated in the eye tracking study. Similar to visual world paradigm experiments involving exposure to native language only, Slavic native speakers have shown to be able to identify a predicate from a non-native, but closely related, language. The outcomes of this study have pointed out that information extracted at the predicate successfully guides eye movements to an object in a visual setting, which satisfies the verb constraints, even if a sentence comes from a non-native language.

Complex experiments involving translation of idiomatic expressions have been designed to understand strategies of providing translation equivalents by Polish and Russian native speakers. The gathered data shows how multi-word expressions tend to be comprehensible for listeners whose L1 is closely related to the stimulus language. However, differences across two types of tasks, that is, open set free translation task, and closed set equivalent selection task, have been observed. Syntactic linear correspondences between perceived phrases, assessed by normalized insertion-deletion measure, have appeared to influence preference of lay translators towards selecting an actual idiomatic equivalent. Overall, it was concluded that phonetic, lexical, and syntactic measures of idiomatic phrase pairs can provide an explanation for strategies of selection of idiomatic equivalents.

The presented studies have pointed out that a degree of intelligibility is highly task-dependent. Particular systems of language show intelligibility effects, which may differ according to research foci. Therefore, it can be concluded that degree of mutual intelligibility depends on a level of linguistic analyses.

8.2 Information-theoretic approach

As shown in this work, intelligibility can be predicted by an information-theoretic notion of surprisal. Stimuli (un)expectedness is often more significant moderator of intelligibility of non-native spoken language than its phonetic resemblance to a native lexicon of a speaker. However, not all the subsystems confirm the accuracy of such a measure. Surprisal has shown to be a good predictor of intercomprehension for selected levels of analyses. The results from the LADO study have pointed out that surprisal-based prediction of the subjects' performance did not reach the threshold of statistical significance. The results of the online language processing study conducted in a short-term priming technique have shown a significant correlation between target stimuli unexpectedness and an increase in latency for all tested languages.

A pilot study involving electromagnetic articulography has been conducted to address the question of surprisal-driven articulation of contextually constrained diphones. The analysis of acoustic and kinematic domain has shown that the degree of coarticulation of the Polish consonant-to-vowel transitions is the highest in the middle surprisal cluster, whereas the low and high surprisal groups are characterized by higher coarticulation resistance and lower spectral overlap. Based on these findings, it has been concluded that surprisal plays an inherent role in motor activity and language production. Furthermore, regardless of small discrepancies across domains, an interplay of information-theoretic and acoustic factors has been discovered. Such a conclusion additionally justified an information-theoretic approach applied in the presented studies regarding the effect of contextual predictability on linguistic encoding.

An information-theoretic approach has been further justified on the basis of the lexical decision task. In a priming study, a notion of surprisal has been an accurate predictor of human performance measured as latency after the exposure to meaningful stimuli coming from other-than-native Slavic languages. The correspondence the higher the surprisal, the longer the latency has been discovered for all four tested languages. This study has provided evidence for context-based word adaptation surprisal outperforming the measures based on phonetic similarities. Therefore, the adapted measure of input (un)expectedness appeared to moderate the performance to a greater extent than cross-lingual phonetic similarity of perceived stimuli. Based on these findings, it has been concluded that quantification of cognate facilitation effect should be supplemented with a surprisal component, and the information-theoretic approach for presenting multilingual lexical relatedness has been validated on the basis of subjects' performance.

Furthermore, on the basis of the eye tracking study, it has been concluded that sentence-level comprehension of closely related languages can be driven by the notion of stimuli (un)expectedness rather than by a resemblance between phonetic surface forms. This outcome has, again, confirmed that more similar-sounding words facilitate intercomprehension to a lower extent than the complexity of sound mappings between languages, which lay listeners may not be aware of. Therefore, is has been concluded that surprisal calculated from regular sound correspondences observed in shared cognates may be able to better distinguish related forms from chance phonetic resemblances, given that unrelated words frequently exhibit chance similarities, and that diachronic processes of sound change do not always produce similar-sounding reflexes of inherited phonemes across related languages.

The study involving translation of idiomatic phrases in Polish and Russian has pointed out that information-theoretic measures can also quantify phrase intelligibility. The applied measure of mean phrase word adaptation surprisal has shown to be a predictor of preference towards literal translation of phrases. In has been observed that the effect of strong phonetic similarities of phrases has motivated the equivalent matching, especially with respect to non-idiomatic equivalents. The results of this experiment have shown the importance of phonetic, lexical, and syntactic cues in the process of equivalent selection for isolated idiomatic phrases.

Various linguistic and extra-linguistic factors influence mutual intelligibility. A non-reciprocal character of intercomprehension presents an additional complexity in quantifying a degree of speech understanding across the investigated language subsystems. Overall, this work has shown the interplay between information-theoretic and linguistic factors in a multilingual environment. Therefore, the methodological apparatus of information theory can be applied in experimental fields of language comprehension studies and multilingualism.

8.3 Asymmetries in intercomprehension

Several presented experiments have shown that the character of Slavic intercomprehension is non-reciprocal. The degree of intercomprehension differs within and across groups of the Slavic languages. Canonical classifications of the Slavic languages do not necessarily reflect actual degree of intercomprehension. For instance, perceptual similarity index, established on the basis of the results of the LADO study, pointed out several discrepancies. Asymmetries observed on the basis of perceptual similarity index reveled that participants placed Polish (with a fixed penultimate word stress) and Czech (with a fixed initial word stress) in different groups, regardless of their typological proximity. Instead, Polish was clustered with Bulgarian, which drives a focus on importance of temporal analyses of perceived vocalic segments. Subjects' performance in the LADO task also confirmed asymmetry beyond classical typology of Slavic languages. It was shown that Bulgarian native speakers more often correctly identified speakers whose L1 was Czech than fellow native speakers of Bulgarian. This finding is related to a dialectological complexity as well as quality of the vocalic segments across tested languages. The results obtained on groups of Bulgarian and Czech native speakers are not in line with a theory that geographical proximity between speech communities and language contact influence intercomprehension.

Furthermore, asymmetry on a lexical level of language processing was discovered. As concluded on the basis of the priming study, the discrepancies among the different language groups suggest asymmetric intelligibility effects in online language processing. What is more, on the basis of the eye tracking study, sentence intelligibility has been shown to depend on comprehenders' native language as well language of stimuli. It was confirmed that the intercomprehension patterns only partly reflect the typological division of the Slavic languages. In accordance with the observed anticipatory eye movements, strong intelligibility effects were discovered for language pairs, which do not belong to the same subgroup of Slavic languages. For instance, the longest latency in object detection was observed in the group of Czech native speakers exposed to Russian stimuli; whereas the exactly the opposite effect, namely, rapid attention shift, was discovered for Russian natives exposed to Czech sentences. A similar effect was observed for Bulgarian native speakers listening to Czech sentences. This study also showed an unidrectional character of comprehension.

A study on comprehension of formulaic expressions pointed out the asymmetry between Polish and Russian native speakers. Interestingly, not only direction of translation but also a character of task introduced differences in providing and matching translation equivalents. In accordance with the gathered dataset, different idiomatic keys have been discovered for Polish and Russian native speakers. It was concluded that information-theoretic measure of the mean phrase word adaptation surprisal is an accurate predictor of subjects' preference towards selecting the literal translation in Polish to Russian direction of translation. However, a lexical distance metrics applied to the translation equivalents showed an opposite pattern in two language groups. On one hand, a strong effect of decrease of lexical distance corresponding to selection of a phrase which contained a cognate lemma was observed in Polish to Russian direction, as opposed to Russian to Polish. On the other hand, a congruent tendency across two groups of native speakers was observed on the level of syntactic analyses of corresponding phrases. It was shown that the greater the syntactic divergence between the target phrases, the stronger the preference for the selection of the idiomatic equivalent.

Typological proximity of tested languages can act in favor of mutual intelligibility among their users. However, language contact and geographical proximity of speech communities does not always reflect a degree of mutual intelligibility. It was shown that, as in a case of Bulgarian and Czech intercomprehension, geographical vicinity seems to play a minor role. Therefore, by referring to a terminology of the pioneering work on receptive multilingualism Voegelin and Harris 1951, a strong chain of mutually intelligible varieties spreads across the Slavic speech communities, regardless of their borders. As shown on the basis of presented experiments, the intercomprehension phenomena only to a limited extend reflect the typological division of the Slavic languages. The intelligibility effects have been discovered for language pairs, which do not belong to the same subgroup of Slavic languages. A tested mode of communication, even when controlled for contextualization cues, often appears to be effortful and depends on linguistic and extra-linguistic factors. Despite observed asymmetries, this work presented evidence for robust mutual intercomprehension among Slavic native speakers, which proved their cognitive linguistic flexibility.

Chapter 9

Conclusions and Outlook

Presented studies showed that spoken intercomprehension among users of the Slavic languages is asymmetric and task-dependent. Phenomena regarding receptive multilingualism do not always adhere to typological divisions of tested languages. The applied experiments have shown that native speakers of Slavic languages exhibit an advantage in comprehension of a spoken utterances coming from a language which does not belong to same grouping. Similar to previous investigations on foreign language comprehension, high task-dependency effect has been confirmed. This work has proposed a theoretical framework, which can be further extended to other language families. Such methodological approach can also find an application in studies on active multilingualism and include various modalities. With an application of hybrid methodology combining experimental phonetics and information theory, it was shown that mutual understanding is often driven by stimuli (un)expectedness rather than phonetic resemblance of perceived stimuli and corresponding unit in a native lexicon.

Based on the LADO study, it was concluded that highly constrained signals can cause an attention shift towards typically less relevant features in spoken language perception. Such features considered vowel quality in the spectral and temporal domains. On the basis of these findings, an engagement of native speakers should be considered in LADO / LOID tests, as well as in forensic procedures. It appeared that speakers of closely related languages can efficiently identify the linguistic origin of a user of another Slavic language. More detailed analyses revealed an importance of subjects' linguistic repertoire in a task of recognition of a language of origin, and pointed out that linguistic background of a listener, even if in command of unrelated languages should not be neglected in similar auditory tasks. Overall, this experiment demonstrated the impressive human capability to identify the origin of a speaker with exposure to highly limited acoustic information.

A study involving analysis of the articulatory gestures, that combined the electromagnetic articulography and acoustic analyses of formant transitions, provided evidence for surprisal-based nature of motor activity even on a diphone level. On the basis of gathered kinematic and acoustic data, it was concluded that coarticulation measured on a set of Polish consonant-to-vowel transitions is a consequence of a twofold surprisal effect. Low coarticulation effects in the low surprisal clusters were understood as need for disambiguation in the absence of the supportive context, whereas such phenomena in the high surprisal groups were explained in the light of prominence given to articulate sequences with low predictability. Therefore, on the basis of the EMA pilot study, it has been concluded that phonotactic predictability can induce changes to speech production patterns. Furthermore, this outcome was interpreted as an argument for an inherent nature of a surprisal component in language production and suggests that the phonotactic unexpectedness of a segment influences motor fluency.

The lexical decision study conducted in a priming paradigm shed light on the interplay of surface interpretation of cognates and non-cognates in associative and phonetic priming. On the basis of the gathered data, it was reported that phonetic word adaptation surprisal and cross-lingual phonetic distance between primes and targets can moderate the latency in an auditory lexical decision task. The phonetic distance seems to impact latency only in Czech and Polish, whereas word adaptation surprisal appears to moderate the response times in all four tested languages. It was concluded that context-based word adaptation surprisal outperforms the measures based on similarity between primes and targets in a short-term priming paradigm. Moreover, the facilitating effect of cognate tokens clustered in the associative priming condition confirms a strong intelligibility effect among the speakers of Bulgarian, Czech, Polish, and Russian. This finding has provided argument for the primacy of associative correspondences and subjects' ability to identify the semantic relatedness of stimuli from another Slavic language without prior linguistic training. Overall, it was concluded that context-based methods for establishing the relation between two meaningful words in closely related languages are better predictors of human performance in a lexical decision task than metrics established exclusively on the basis of the similarity of stimuli.

In the eye tracking study, a comprehension of spoken stimuli has been tested in a visual environment. This experiment put a focus on sentence-level comprehension. The data gathered in this study showed that Slavic native speakers can immediately establish dependency between the predicate of a sentence and its direct object, even if a spoken stimuli comes from their non-native language. The collected data supported the hypothesis that sentence processing, measured by anticipatory eye movements, was driven by the information-theoretic notion of surprisal based on corresponding predicates. This experiment also showed an asymmetric character of intercomprehension across the four groups of native speakers. This observation is in line with previous studies reporting asymmetrical character of intercomprehension. The experiment has demonstrated that information extracted at the predicate successfully guides eye movements to a visual object which satisfies the verb constraints. Such gaze anticipation have also been moderated by a degree of stimuli (un)expectedness, which supports an argument for surprisal-driven intelligibility effect among Slavic native speakers.

The study concerning translation of idiomatic phrases showed that strong surface phonetic similarities of phrases seem to motivate the equivalent matching, especially with respect to literal translation equivalents. Overall, the gathered data revealed that subjects' preference to choose a particular type of translation equivalent can be predicted by phonetic, lexical, and syntactic similarities. However, an interesting asymmetry across tested languages was observed and differences between open set and closed set task were noticed. The experiment showed that the nature of mutual intelligibility is unidirectional, and different idiomatic keys have been discovered across the groups of participants. Even though comprehension of idioms presents a difficult task, the gathered data showed how idiomatic expressions tend to be comprehensible for listeners whose L1 is closely related to the stimulus language. Overall, the data suggested that phonetic, lexical, and syntactic measures of pairs of idiomatic expressions can provide an explanation for strategies used by native speakers of closely related languages in the selection of translation equivalents.

Further studies on receptive multilingualism of speakers of closely related languages could have been focused on other language families with various degrees of typological proximity. Another interesting perspective to research on communication of interactants coming from various speech communities can lead to a framework of areal linguistics. In this case, languages in contact rather than genetically close vernaculars could have been taken into consideration. Proposed methodological approach could also be applied to studies concerning language acquisition and attrition in which contextual quantification of linguistic stimuli can be performed with an adaptation of information-theoretic apparatus. Additionally, future investigation could have included multimodal tasks, which would develop an experimental paradigm for discovering cross-lingual and cross-cultural correspondences contributing to mutual understanding.

List of Tables

List of Figures

Bibliography

Abdullah, B. M., T. Avgustinova, B. Möbius, and D. Klakow (2020). "Cross-Domain Adaptation of Spoken Language Identification for Related Languages: The Curious Case of Slavic Languages". In: *arXiv preprint arXiv:2008.00545*. URL: https://arxiv.org/abs/2008.00545.

Akaike, H. (1998a). "Information theory and an extension of the maximum likelihood principle". In: *Selected Papers of Hirotugu Akaike*. Ed. by E. Parzen, K. Tanabe, and G. Kitagawa. Springer, pp. 199–213.

Akaike, H. (1998b). "Prediction and entropy". In: *Selected Papers of Hirotugu Akaike*. Ed. by E. Parzen, K. Tanabe, and G. Kitagawa. Springer, pp. 387–410.

Allopenna, P. D., J. S. Magnuson, and M. K. Tanenhaus (1998). "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models". In: *Journal of Memory and Language* 38.4, pp. 419–439. DOI: 10.1006/jmla.1997.2558.

Altmann, G. T. M. and Y. Kamide (1999). "Incremental interpretation at verbs: Restricting the domain of subsequent reference". In: *Cognition* 73.3, pp. 247–264. DOI: 10.1016/S0010-0277(99)00059-1.

Aylett, M. and A. Turk (2004). "The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech". In: *Language and Speech* 47.1, pp. 31–56. DOI: 10.1177/00238309040470010201.

Aylett, M. and A. Turk (2006). "Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei". In: *The Journal of the Acoustical Society of America* 119.5, pp. 3048–3058. DOI: 10.1121/1.2188331.

Bakran, J. and V. Mildner (1995). "Effect of speech rate and coarticulation strategies on the locus equation determination". In: *Proceedings of the XIIIth International Congress of Phonetic Sciences*. Vol. 1, pp. 26–29.

Bartula, C. (2011). *Podstawowe wiadomości z gramatyki staro-cerkiewno-słowiańskiej na tle porównawczym*. Warszawa, Państwowe Wydawnictwo Naukowe.

Beerkens, R. (2010). *Receptive multilingualism as a language mode in the Dutch-German border area*. Waxmann Verlag.

Bell, A., J. M. Brenier, M. Gregory, C. Girand, and D. Jurafsky (2009). "Predictability effects on durations of content and function words in conversational English". In: *Journal of Memory and Language* 60.1, pp. 92–111. DOI: 10.1016/j.jml.2008.06.003.

Berman, A. (2000). "Translation and the Trials of the Foreign". In: *The Translation Studies Reader*. Ed. by L. Venuti. London, Routledge, pp. 284–297.

Berry, J. J. (2011). "Accuracy of the NDI wave speech research system". In: *Journal of Speech, Language, and Hearing Research* 54.5, pp. 1295–1301. DOI: 10.1044/1092-4388(2011/10-0226).

Bezooijen, R. van and C. Gooskens (2007). "Interlingual text comprehension: Linguistic and extralinguistic determinants". In: *Receptive Multilingualism: Linguistic analyses, language policies, and didactic concepts*. Ed. by J. D. ten Thije and L. Zeevaert. John Benjamins, pp. 249–263.

Bobrow, S. A. and S. M. Bell (1973). "On catching on to idiomatic expressions". In: *Memory & Cognition* 1.3, pp. 343–346. DOI: 10.3758/BF03198118.

Boers, F. and M. Demecheleer (2001). "Measuring the impact of cross-cultural differences on learners' comprehension of imageable idioms". In: *ELT Journal* 55.3, pp. 255–262. DOI: 10.1093/elt/55.3.255.

Boersma, P. and D. Weenink (2020). "Praat: Doing phonetics by computer. Version 6.1.16". In: URL: https://www.fon.hum.uva.nl/praat/.

Bonatti, L. L., M. Pena, M. Nespor, and J. Mehler (2005). "Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing". In: *Psychological Science* 16.6, pp. 451–459. DOI: 10.1111/j.0956-7976.2005.01556.x.

Bonvino, E. (2015). "Intercomprehension studies in Europe: history, current methodology and future developments". In: *Intercomprehension and Plurilingualism: An Asset for Italian in the USA*. Ed. by R.

Dolci and A. Tamburri. John D. Calandra Italian American Institute Queens College, CUNY, pp. 29–59.

Brandt, E. (2019). *Information density and phonetic structure: Explaining segmental variability*. Saarländische Universitäts- und Landesbibliothek. DOI: doi:10.22028/D291-27918.

Braunmüller, K. (2007). "Receptive multilingualism in northern Europe in the Middle Ages: A description of a scenario". In: *Receptive multilingualism: Linguistic analyses, language policies and didactic concepts*. Ed. by J. D. ten Thije and L. Zeevaert. John Benjamins, pp. 25–47.

Braunmüller, K. (2008). *On the relevance of receptive multilingualism in a globalised world: theory, history and evidence from today's Scandinavia*. Universität Hamburg.

Bray, J. H. and S. E. Maxwell (1985). "Multivariate analysis of variance". In: *Quantitative Applications in the Social Sciences*. Vol. 54. Sage. DOI: 10.4135/9781412985222.

Bugarski, R. (2018). *Govorite li zajednički? Kako je nastala i kako je primljena Deklaracija o zajedničkom jeziku*. Beograd, Biblioteka XX vek.

Bulatović, S., A. Schüppert, and C. Gooskens (2019). "Receptive multilingualism versus ELF: How well do Slovenes understand Croatian compared to Croatian speakers' English?" In: *Journal of English as a Lingua Franca* 8.1, pp. 37–65. DOI: 10.1515/jelf-2019-2005.

Cacciari, C. and P. Tabossi (1988). "The comprehension of idioms". In: *Journal of Memory and Language* 27.6, pp. 668–683. DOI: 10.1016/0749-596X(88)90014-9.

Cambier-Langeveld, T. (2016). "Language analysis in the asylum procedure: a specification of the task in practice." In: *International Journal of Speech, Language & the Law* 23.1, pp. 25–41. DOI: 10.1558/ijsll.v23i1.17539.

Carter, B. T. and S. G. Luke (2020). "Best practices in eye tracking research". In: *International Journal of Psychophysiology* 155, pp. 49–62. DOI: 10.1016/j.ijpsycho.2020.05.010.

Casad, E. (1974). "Dialect intelligibility testing". In: *SIL Publications in Linguistics and Related Fields*. Ed. by I. Davis. Vol. 38. Summer Institute of Linguistics.

Chlebda, W. M. (2016). *Podręczny idiomatykon polsko-rosyjski. Z. 8.* Wydawnictwo Uniwersytetu Opolskiego.

Cieślicka, A. B. (2015). "Idiom acquisition and processing by second/foreign language learners." In: *Bilingual figurative language processing.* Ed. by R. R. Heredia and A. B. Cieślicka. Cambridge University Press, pp. 208–244. DOI: 10.1017/CBO9781139342100.012.

Cieślicka, A. B., R. R. Heredia, and T. Garcıa (2017). "Task effects in bilingual idiom comprehension". In: *Poznan Studies in Contemporary Linguistics* 53.1, pp. 95–117. DOI: 10.1515/psicl-2017-0005.

Cooper, R. M. (1974). "The control of eye fixation by the meaning of spoken language: a new methodology for the real-time investigation of speech perception, memory, and language processing." In: *Cognitive Psychology* 6.1. DOI: 10.1016/0010-0285(74)90005-X.

Costa, A., A. Caramazza, and N. Sebastian-Galles (2000). "The cognate facilitation effect: implications for models of lexical access." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26.5, pp. 1283–1296. DOI: 10.1037/0278-7393.26.5.1283.

Crocker, M. W., V. Demberg, and E. Teich (2016). "Information density and linguistic encoding (IDeaL)". In: *KI-Künstliche Intelligenz* 30.1, pp. 77–81. DOI: 10.1007/s13218-015-0391-y.

Cronk, B. C. and W. A. Schweigert (1992). "The comprehension of idioms: The effects of familiarity, literalness, and usage". In: *Applied Psycholinguistics* 13.2, pp. 131–146. DOI: 10 . 1017 / S0142716400005531.

Cutler, A. (1982). "Idioms: The colder the older". In: *Linguistic Inquiry* 13.2, pp. 317–320.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words.* MIT Press.

Dalewska-Greń, H. (2007). *Języki słowiańskie.* Warszawa, PWN.

Delogu, F., M. W. Crocker, and H. Drenhaus (2017). "Teasing apart coercion and surprisal: Evidence from eye-movements and ERPs". In: *Cognition* 161, pp. 46–59. DOI: 10.1016/j.cognition.2016.12.017.

Demberg, V., A. Sayeed, P. Gorinski, and N. Engonopoulos (2012). "Syntactic surprisal affects spoken word duration in conversational contexts". In: *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, pp. 356–367. URL: https://aclanthology.org/D12-1000.

Dijkstra, T. (2005). "Bilingual visual word recognition and lexical access". In: *Handbook of bilingualism: Psycholinguistic approaches.* Ed. by J. F. Kroll and A. M. B. de Groot. Oxford University Press, pp. 197–201.

Dijkstra, T., B. Hilberink-Schulpen, and W. J. B. van Heuven (2010). "Repetition and masked form priming within and between languages using word and nonword neighbors". In: *Bilingualism: Language and Cognition* 13, pp. 341–357. DOI: 10.1017/S1366728909990575.

Doetjes, G. (2007). "Understanding differences in inter-Scandinavian language understanding". In: *Receptive multilingualism. Linguistic analyses, language policies and didactic concepts*. Ed. by J. D. ten Thije and L. Zeevaert. Benjamins Amsterdam, pp. 217–230.

Donato, C. and V. Pasquarelli-Gascon (2015). "The language of the other: Italian for Spanish speakers through intercomprehension". In: *Italica* 92.3, pp. 713–735.

Dromey, C., E. Hunter, and S. L. Nissen (2018). "Speech adaptation to kinematic recording sensors: Perceptual and acoustic findings". In: *Journal of Speech, Language, and Hearing Research* 61.3, pp. 593–603. DOI: 10.1044/2017_JSLHR-S-17-0169.

Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.

Dulichenko, A. D. (1981). *Slavjanskie literaturnye mikrojazyki: voprosy formirovanija i razvitija*. Valgus.

Dunn, O. J. (1964). "Multiple comparisons using rank sums". In: *Technometrics* 6.3, pp. 241–252.

Duyck, W. (2005). "Translation and associative priming with cross-lingual pseudohomophones: evidence for nonselective phonological activation in bilinguals." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31.6, p. 1340. DOI: 10.1037/0278-7393.31.6.1340.

Eades, D. (2005). "Applied linguistics and language analysis in asylum seeker cases". In: *Applied Linguistics* 26.4, pp. 503–526. DOI: 10.1093/applin/ami021.

Engstrand, O. and B. Lindblom (1997). "The locus line: does aspiration affect its steepness?" In: *Fonetik*. Vol. 97, pp. 101–104.

European Communities, C. of the (2007). *High Level Group on Multilingualism: final report*. Office for Official Publications of the European Communities Luxembourg. URL: http://hdl.voced.edu.au/10707/75636.

Farris-Trimble, A. and B. McMurray (2013). "Test–retest reliability of eye tracking in the visual world paradigm for the study of real-time spoken word recognition". In: *Journal of Speech, Language, and Hearing Research* 56.4. DOI: 10.1044/1092-4388(2012/12-0145).

Fedorov, A. I. (1995). *Frazeologičeskij slovar russkogo literaturnogo jazyka. 2. N-Ja.* Nauka.

Fougeron, C. and P. A. Keating (1997). “Articulatory strengthening at edges of prosodic domains”. In: *The Journal of the Acoustical Society of America* 101.6, pp. 3728–3740. DOI: 10.1121/1.418332.

Foulkes, P. and K. Wilson (2011). “Language Analysis for the Determination of Origin: An Empirical Study”. In: *Proc. ICPhS 2011*, pp. 691–694.

Fowler, C. A. (1994). “Invariants, specifiers, cues: An investigation of locus equations as information for place of articulation”. In: *Perception & Psychophysics* 55.6, pp. 597–610. DOI: 10.3758/BF03211675.

Fraser, B. (1970). “Idioms within a transformational grammar”. In: *Foundations of Language* 6, pp. 22–42.

Fraser, H. (2009). “The role of ’educated native speakers’ in providing language analysis for the determination of the origin of asylum seekers.” In: *International Journal of Speech, Language & the Law* 16.1. DOI: 10.1558/ijsll.v16i1.113.

Fraser, H. (2011). “The role of linguists and native speakers in language analysis for the determination of speaker origin”. In: *International Journal of Speech, Language & the Law* 18.1. DOI: 10.1558/ijsll.v18i1.121.

Fruchter, D. and H. M. Sussman (1997). “The perceptual relevance of locus equations”. In: *The Journal of the Acoustical Society of America* 102.5, pp. 2997–3008. DOI: 10.1121/1.421012.

Gibbs, R. W. (1980). “Spilling the beans on understanding and memory for idioms in conversation”. In: *Memory & Cognition* 8.2, pp. 149–156. DOI: 10.3758/BF03213418.

Golubović, J. (2016). *Mutual intelligibility in the Slavic language area.* Groningen, Center for Language and Cognition.

Golubović, J. and C. Gooskens (2015). “Mutual intelligibility between West and South Slavic languages”. In: *Russian Linguistics* 39.3, pp. 351–373. DOI: 10.1007/s11185-015-9150-9.

Gooskens, C. (2007). “The contribution of linguistic factors to the intelligibility of closely related languages”. In: *Journal of Multilingual and Multicultural Development* 28.6, pp. 445–467. DOI: 10.2167/jmmd511.0.

Gooskens, C. (2013). “Experimental methods for measuring intelligibility of closely related language varieties”. In: *The Oxford Handbook of Sociolinguistics.* Ed. by R. Bayley, R. Cameron, and C. Lucas. Oxford University Press, pp. 195–213.

Gooskens, C. (2019). “Receptive multilingualism”. In: *Multidisciplinary Perspectives on Multilingualism*. Ed. by S. Montanari and S. Quay. Language Contact and Bilingualism. De Gruyter Mouton, pp. 149–174. DOI: 10.1515/9781501507984-008.

Gooskens, C. and R. van Bezooijen (2006). “Mutual comprehensibility of written Afrikaans and Dutch: Symmetrical or asymmetrical?” In: *Literary and Linguistic Computing* 21.4, pp. 543–557. DOI: 10.1093/llc/fql036.

Gooskens, C. and V. J. van Heuven (2017). “Measuring cross-linguistic intelligibility in the Germanic, Romance and Slavic language groups”. In: *Speech Communication* 89, pp. 25–36. DOI: 10.1016/j.specom.2017.02.008.

Gooskens, C. and V. J. van Heuven (2020). “How well can intelligibility of closely related languages in Europe be predicted by linguistic and non-linguistic variables?” In: *Linguistic Approaches to Bilingualism* 10.3, pp. 351–379. DOI: 10.1075/lab.17084.goo.

Gooskens, C. and V. J. van Heuven (2021). “Mutual intelligibility”. In: *Similar languages, varieties, and dialects: A computational perspective*. Ed. by M. Zamperi and P. Nakov. Cambridge University Press, pp. 51–95.

Gooskens, C., V. J. van Heuven, J. Golubović, A. Schüppert, F. Swarte, and S. Voigt (2018). “Mutual intelligibility between closely related languages in Europe”. In: *International Journal of Multilingualism* 15.2, pp. 169–193. DOI: 10.1080/14790718.2017.1350185.

Gooskens, C. and F. Swarte (2017). “Linguistic and extra-linguistic predictors of mutual intelligibility between Germanic languages”. In: *Nordic Journal of Linguistics* 40.2, pp. 123–147. DOI: 10.1017/S0332586517000099.

Gor, K. (2018). “Phonological priming and the role of phonology in non-native word recognition”. In: *Bilingualism* 21.3, pp. 437–442. DOI: 10.1017/S1366728918000056.

Grice, H. P. (1975). “Logic and conversation”. In: *Speech acts*. Ed. by P. Cole and J. L. Morgan. Brill, pp. 41–58. DOI: 10.1163/9789004368811_003.

Hale, J. (2001). “A Probabilistic Earley Parser as a Psycholinguistic Model”. In: *Second meeting of the north American chapter of the association for computational linguistics*. URL: https://aclanthology.org/N01-1021.pdf.

Hale, J. (2016). “Information-theoretical complexity metrics”. In: *Language and Linguistics Compass* 10.9, pp. 397–412. DOI: 10.1111/lnc3.12196.

Harper, M. and M. Maxwell (2008). “Spoken language characterization”. In: *Springer Handbook of Speech Processing*. Ed. by M. S. J. Benesty and Y. Huang. Springer, pp. 797–810. DOI: 10.1007/978-3-540-49127-9_40.

Haugen, E. (1966). “Semicommunication: The language gap in Scandinavia”. In: *Sociological Inquiry* 36.2, pp. 280–297. DOI: 10.1111/j.1475-682X.1966.tb00630.x.

Hay, J., P. Warren, and K. Drager (2006). “Factors influencing speech perception in the context of a merger-in-progress”. In: *Journal of Phonetics* 34.4, pp. 458–484. DOI: 10.1016/j.wocn.2005.10.001.

Heeringa, W., F. Swarte, A. Schüppert, and C. Gooskens (2018). “Measuring syntactical variation in Germanic texts”. In: *Digital Scholarship in the Humanities* 33.2, pp. 279–296. DOI: 10.1093/llc/fqx029.

Hejwowski, K. (2004). *Kognitywno-komunikacyjna teoria przekładu.* Warszawa, Wydawnictwo Naukowe PWN.

Hickerson, H. and N. Hickerson (1952). “Testing procedures for estimating transfer of information among Iroquois dialects and languages”. In: *International Journal of American Linguistics* 18.1, pp. 1–8. DOI: 10.1086/464140.

Hoskin, J. (2018). “Native speaker non-linguists in LADO: an insider perspective”. In: *Forensic Linguistics: Asylum-seekers, Refugees and Immigrants.* Ed. by I. M. Nick. Vernon Press, pp. 23–40.

House, A. S. and E. P. Neuburg (1977). “Toward automatic identification of the language of an utterance. I. Preliminary methodological considerations”. In: *The Journal of the Acoustical Society of America* 62.3, pp. 708–713. DOI: 10.1121/1.381582.

Hume, E. (2016). “Phonological markedness and its relation to the uncertainty of words”. In: *On-in Kenkyu [Phonological Studies]* 19, pp. 107–116.

Iskarous, K., C. A. Fowler, and D. H. Whalen (2010). “Locus equations are an acoustic expression of articulator synergy”. In: *The Journal of the Acoustical Society of America* 128.4, pp. 2021–2032. DOI: 10.1121/1.3479538.

Ivić, P. (2001). *Dijalektologija srpskohrvatskog jezika: Uvod u štokavsko narečje.* Vol. 2. Sremski Karlovci - Novi Sad, Izdavačka knjižarnica Zorana Stojanovića.

Jadlovsk, T. (2020). "Translation of Non-Equivalent Idioms (using English and Russian examples)". In: *Armenian Folia Anglistika* 16.2 (22), pp. 90–105. DOI: 10.46991/AFA/2020.16.2.090.

Jaeger, T. F. (2010). "Redundancy and reduction: Speakers manage syntactic information density". In: *Cognitive Psychology* 61.1, pp. 23–62. DOI: 10.1016/j.cogpsych.2010.02.002.

Jágrová, K., T. Avgustinova, I. Stenger, and A. Fischer (2019). "Language models, surprisal and fantasy in slavic intercomprehension". In: *Computer Speech & Language* 53, pp. 242–275. DOI: 10.1016/j.csl.2018.04.005.

Jensen, J. B. (1989). "On the mutual intelligibility of Spanish and Portuguese". In: *Hispania* 72.4, pp. 848–852.

Ji, A., J. J. Berry, and M. T. Johnson (2014). "The Electromagnetic Articulography Mandarin Accented English (EMA-MAE) corpus of acoustic and 3D articulatory kinematic data". In: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 7719–7723. DOI: 10.1109/ICASSP.2014.6855102.

Johnson, K. (2003). *Acoustic and auditory phonetics*. 2nd ed. Maiden, Blackwell Publishing.

Joordens, S. and S. Becker (1997). "The long and short of semantic priming effects in lexical decision." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23.5, pp. 1083–1105. DOI: 10.1037/0278-7393.23.5.1083.

Jurafsky, D., A. Bell, M. Gregory, and W. D. Raymond (2001). "Probabilistic relations between words: Evidence from reduction in lexical production". In: *Frequency and the Emergence of Linguistic Structure*. Ed. by J. L. Bybee and P. L. Hopper. Vol. 45, pp. 229–254.

Kamide, Y., G. T. M. Altmann, and S. L. Haywood (2003). "The timecourse of prediction in incremental sentence processing: Evidence from anticipatory eye movements". In: *Journal of Memory and Language* 49.1, pp. 133–156. DOI: 10.1016/S0749-596X(03)00023-8.

Kelley, M. C. and B. V. Tucker (2020). "A comparison of four vowel overlap measures". In: *The Journal of the Acoustical Society of America* 147.1, pp. 137–145. DOI: 10.1121/10.0000494.

Kessler, R., A. Weber, and C. K. Friedrich (2020). "Activation of Literal Word Meanings in Idioms: Evidence from Eye-tracking and ERP Experiments". In: *Language and Speech*, pp. 594–624. DOI: 10.1177/0023830920943625.

Kirchhoff, K. and S. Parandekar (2001). "Multi-stream statistical N-gram modeling with application to automatic language identification". In: *Proc. Eurospeech 2001.* URL: https://www.isca-speech.org/archive_v0/archive_papers/eurospeech_2001/e01_0803.pdf.

Kisler, T., U. Reichel, and F. Schiel (2017). "Multilingual processing of speech via web services". In: *Computer Speech & Language* 45, pp. 326–347. DOI: 10.1016/j.csl.2017.01.005.

Kordić, S. (2010). *Jezik i nacionalizam.* Zagreb, Durieux.

Kovecses, Z. and P. Szabco (1996). "Idioms: A view from cognitive semantics". In: *Applied Linguistics* 17.3, pp. 326–355. DOI: 10.1093/applin/17.3.326.

Kowalski, P. (2015). "Współczesny literacki język prekmurski (casus Porabja)". In: *Poznańskie Studia Slawistyczne* 8. DOI: 10.14746/pss.2015.8.6.

Krippendorff, K., S. W. Littlejohn, and K. A. Foss (2009). "Mathematical Theory of Communication". In: *Encyclopedia of communication theory.* Ed. by S. W. Littlejohn and K. A. Foss. Sage, pp. 614–618.

Krull, D. (1989). "Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech". In: *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm.* Ed. by O. Engstrand, M. Dufberg, and C. Kylander. Vol. 10, pp. 87–108.

Kudera, J., P. Georgis, H. M. T. Alam, B. Möbius, T. Avgustinova, and D. Klakow (2022). "Comprehension of closely related languages: A visual world eye tracking study". In: *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2022.* Ed. by O. Niebuhr, M. S. Lundmark, and H. Weston. TUDpress, Dresden, pp. 212–219.

Kudera, J., P. Georgis, B. Möbius, T. Avgustinova, and D. Klakow (2021a). "Phonetic Distance and Surprisal in Multilingual Priming: Evidence from Slavic". In: *Proc. Interspeech 2021*, pp. 3944–3948. DOI: 10.21437/Interspeech.2021-1003.

Kudera, J., L. Tavi, B. Möbius, T. Avgustinova, and D. Klakow (2021b). "The Effect of Surprisal on Articulatory Gestures in Polish Consonant-to-Vowel Transitions: A Pilot EMA Study". In: *ITG-Fachbericht 298: Speech Communication.* VDE Verlag, pp. 179–183. URL: https://ieeexplore.ieee.org/document/9657527.

Kürschner, S., C. Gooskens, and R. van Bezooijen (2008). "Linguistic determinants of the intelligibility of Swedish words among Danes". In: *International Journal of Humanities and Arts Computing* 2.1-2, pp. 83–100. DOI: 10.3366/E1753854809000329.

Kuryłowicz, J. (1957). “O jedności językowej bałto-słowiańskiej”. In: *Biuletyn Polskiego Towarzystwa Językoznawczego* XVI, pp. 71–113.

Ladefoged, P. and I. Maddieson (1996). *The sounds of the world's languages*. Oxford, Blackwell.

Lalor, E. and K. Kirsner (2000). “Cross-lingual transfer effects between English and Italian cognates and noncognates”. In: *International Journal of Bilingualism* 4.3, pp. 385–398. DOI: 10.1177/13670069000040030501.

Lambelet, A. and P.-Y. Mauron (2017). “Receptive multilingualism at school: an uneven playing ground?” In: *International Journal of Bilingual Education and Bilingualism* 20.7, pp. 854–867. DOI: 10.1080/13670050.2015.1114583.

LANG (2004). “Guidelines for the use of language analysis in relation to questions of national origin in refugee cases”. In: *International Journal of Speech, Language and the Law* 11.2, pp. 261–266. DOI: 10.1558/ijsll.v11i2.261.

Lee, W.-S. (2017). “Co-Articulation between Consonant and Vowel in Cantonese Syllables”. In: *International Journal of Cognitive and Language Sciences* 11.2, pp. 417–422.

Lehr-Spławiński, T. (1957). “O tzw. przestawce płynnych w językach słowiańskich”. In: *Studia i szkice wybrane z językoznawstwa słowiańskiego*. Warszawa, Państwowe Wydawnictwo Naukowe, pp. 231–250.

Levenshtein, V. (1965). “Binary Codes Capable of Correcting Deletions, Insertions and Reversals”. In: *Doklady Akademii Nauk SSSR* 163.4, pp. 845–848.

Levy, R. (2008). “Expectation-based syntactic comprehension”. In: *Cognition* 106.3, pp. 1126–1177. DOI: 10.1016/j.cognition.2007.05.006.

Lieberman, P. (1963). “Some effects of semantic and grammatical context on the production and perception of speech”. In: *Language and Speech* 6.3, pp. 172–187. DOI: 10.1177/002383096300600306.

Lindblom, B. (1963). *On vowel reduction*. Stockholm, Royal Institute of Technology.

Löfqvist, A. (1999). “Interarticulator phasing, locus equations, and degree of coarticulation”. In: *The Journal of the Acoustical Society of America* 106.4, pp. 2022–2030. DOI: 10.1121/1.427948.

Luce, P. A., S. D. Goldinger, E. T. Auer, and M. S. Vitevitch (2000). “Phonetic priming, neighborhood activation, and PARSYN”. In:

Perception & Psychophysics 62.3, pp. 615–625. DOI: 10 . 3758 / BF03212113.

Lukszyn, J., ed. (1998). *Wielki słownik frazeologiczny polsko-rosyjski, rosyjsko-polski.* Warszawa.

Mareš, F. V. (1969). “Diachronische Phonologie des Ur-und Frühslavischen: Slavistische Beiträge”. In: *Slavistische Beiträge.* Vol. 40. München, Sagner.

Matasović, R. (2005). “Toward a relative chronology of the earliest Baltic and Slavic sound changes”. In: *Baltistica* 40.2, pp. 147–157.

McQueen, J. M., D. Dahan, and A. Cutler (2011). “Continuity and gradedness in speech processing”. In: *Phonetics and Phonology in Language Comprehension and Production.* Ed. by N. O. Schiller and A. S. Meyer. De Gruyter Mouton, pp. 39–78.

Meissner, F.-J. (2018). “Modelling Plurilingual Processing in Intercomprehension”. In: *Intercompréhension en réseau. Scénarios, médiations, évaluations.* Ed. by S. Garbarino and C. Degache, pp. 279–283.

Melo-Pfeifer, S. (2014). “Intercomprehension between Romance Languages and the role of English: a study of multilingual chat rooms”. In: *International Journal of Multilingualism* 11.1, pp. 120–137. DOI: 10.1080/14790718.2012.679276.

Ministerådet, N. (2006). *Deklaration om nordisk språkpolitik.* URL: http://norden.diva-portal.org/smash/get/diva2:700895/FULLTEXT01.pdf.

Moberg, J., C. Gooskens, J. Nerbonne, and N. Vaillette (2007). “Conditional Entropy Measures Intelligibility among Related Languages”. In: *Proceedings of Computational Linguistics in the Netherlands.* Ed. by F. van Eynde, P. Dirix, I. Schuurman, and V. Vandeghinste, pp. 51–66.

Modarresi, G., H. M. Sussman, B. Lindblom, and E. Burlingame (2005). “Locus equation encoding of stop place: revisiting the voicing/VOT issue”. In: *Journal of Phonetics* 33.1, pp. 101–113. DOI: 10.1016/j.wocn.2004.06.002.

Moon, S.-J. and B. Lindblom (1994). “Interaction between duration, context, and speaking style in English stressed vowels”. In: *The Journal of the Acoustical Society of America* 96.1, pp. 40–55. DOI: 10.1121/1.410492.

Mosbach, M., I. Stenger, T. Avgustinova, and D. Klakow (2019). "incom.py - A Toolbox for Calculating Linguistic Distances and Asymmetries between Related Languages". In: *Proceedings of Recent Advances in Natural Language Processing, RANLP 2019*. Varna, Bulgaria, pp. 811–819. DOI: 10.26615/978-954-452-056-4_094. published.

Moszyński, L. (2006). *Wstęp do filologii słowiańskiej*. 2nd ed. Warszawa, Wydawnictwo Naukowe PWN.

Muthusamy, Y. K., N. Jain, and R. A. Cole (1994). "Perceptual benchmarks for automatic language identification". In: *Proceedings of ICASSP'94. IEEE International Conference on Acoustics, Speech and Signal Processing*. Vol. 1. IEEE, pp. I/333–I/336. DOI: 10.1109/ICASSP.1994.389288.

Navratil, J. (2001). "Spoken language recognition-a step toward multilinguality in speech processing". In: *IEEE Transactions on Speech and Audio Processing* 9.6, pp. 678–685. DOI: 10.1109/89.943345.

Needleman, S. B. and C. D. Wunsch (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins". In: *Journal of Molecular Biology* 48.3, pp. 443–453. DOI: 10.1016/0022-2836(70)90057-4.

Nicol, J. L. (1996). "Syntactic priming". In: *Language and Cognitive Processes* 11.6, pp. 675–680. DOI: 10.1080/016909696387088.

Nolan, F. (2012). "Degrees of freedom in speech production: an argument for native speakers in LADO." In: *International Journal of Speech, Language & the Law* 19.2. DOI: 10.1558/ijsll.v19i2.263.

Nunberg, G., I. A. Sag, and T. Wasow (1994). "Idioms". In: *Language* 70.3, pp. 491–538. DOI: 10.1353/lan.1994.0007.

Nycz, J. and L. Hall-Lew (2013). "Best practices in measuring vowel merger". In: *Proceedings of Meetings on Acoustics* 20.1, pp. 1–19. DOI: 10.1121/1.4894063.

Ogle, D. H., P. Wheeler, and A. Dinno (2021). *FSA: Fisheries Stock Analysis*. R package version 0.8.32. URL: https://github.com/droglenc/FSA.

Ohala, J. J. and J. B. Gilbert (1979). "Listeners' ability to identify languages by their prosody". In: *Problèmes de Prosodie*. Ed. by P. Léon and M. Rossi, pp. 123–131.

Papoutsaki, A., P. Sangkloy, J. Laskey, N. Daskalova, J. Huang, and J. Hays (2016). "WebGazer: Scalable Webcam Eye Tracking Using User Interactions". In: *Proceedings of the 25th International Joint*

Conference on Artificial Intelligence (IJCAI). AAAI, pp. 3839–3845. URL: https://www.ijcai.org/Proceedings/16/Papers/540.pdf.

Patrick, P. L. (2010). "Language variation and LADO (language analysis for determination of origin)". In: *Language and Origin. The Role of Language in European Asylum Procedures: A Linguistic and Legal Perspectives*. Ed. by K. Zwaan, M. Verrips, and P. Muysken. Wolf Legal Publishers, pp. 73–87.

Patrick, P. L. (2012). "Language Analysis For Determination Of Origin: Objective Evidence For Refugee Status Determination". In: *The Oxford Handbook of Language and Law*. Ed. by L. M. Solan and P. M. Tiersma, pp. 533–546. DOI: 10.1093/oxfordhb/9780199572120.013.0039.

Peeters, D., T. Dijkstra, and J. Grainger (2013). "The representation and processing of identical cognates by late bilinguals: RT and ERP effects". In: *Journal of Memory and Language* 68.4, pp. 315–332. DOI: https://doi.org/10.1016/j.jml.2012.12.003.

Pellegrino, F. and R. André-Obrecht (2000). "Automatic language identification: an alternative approach to phonetic modelling". In: *Signal Processing* 80.7, pp. 1231–1244. DOI: 10.1016/S0165-1684(00)00032-3.

Peperkamp, S. and E. Dupoux (2002). "A typological study of stress 'deafness'". In: *Laboratory phonology 7*. Ed. by C. Gussenhoven and N. Warner. De Gruyter Mouton, pp. 203–240. DOI: 10.1515/9783110197105.1.203.

Peperkamp, S., I. Vendelin, and E. Dupoux (2010). "Perception of predictable stress: A cross-linguistic investigation". In: *Journal of Phonetics* 38.3, pp. 422–430. DOI: 10.1016/j.wocn.2010.04.001.

Perea, M. and E. Rosa (2002). "The effects of associative and semantic priming in the lexical decision task". In: *Psychological Research* 66, pp. 180–194. DOI: 10.1007/s00426-002-0086-5.

Perillo, S., H.-Y. Bang, and M. Clayards (2017). "Locus equation metrics as an index of coarticulation resistance: The effect of prosodic prominence". In: *Proceedings of Meetings on Acoustics* 25.1. DOI: 10.1121/2.0000390.

Petrova, T. E. and E. Mikhailovskaya (2014). "Processing of idioms: Eye-tracking study on Russian language material". In: *Global Scientific Potential* 10.43, pp. 160–162.

Pęzik, P. (2012). "Wyszukiwarka PELCRA dla danych NKJP". In: *Narodowy korpus języka polskiego*. Ed. by A. Przepiórkowski, M. Bańko,

R. Górski, and B. Lewandowska-Tomaszczyk. Warszawa, PWN, pp. 253–274.

Piantadosi, S. T., H. Tily, and E. Gibson (2011). "Word lengths are optimized for efficient communication". In: *Proceedings of the National Academy of Sciences* 108.9, pp. 3526–3529. DOI: 10.1073/pnas.1012551108.

Pillai, K. C. S. (1954). "On some distribution problems in multivariate analysis". In: *Mimeographed Series.* 88. Institute of Statistics, University of North Carolina, Chapel Hill, North Carolina. URL: https://repository.lib.ncsu.edu/bitstream/handle/1840.4/2164/ISMS_1954_88.pdf?sequence=1.

Ploeg, M. van der, F. Swarte, and C. Gooskens (2017). "The effect of age and level of education on intelligibility". In: *Tijdschrift voor Skandinavistiek* 35.1, pp. 56–70.

Pluymaekers, M., M. Ernestus, and R. H. Baayen (2005). "Articulatory Planning Is Continuous and Sensitive to Informational Redundancy". In: *Phonetica* 62.2-4, pp. 146–159. DOI: 10.1159/000090095.

Popowska-Taborska, H. (1984). *Z dawnych podziałów Słowiańszczyzny: słowiańska alternacja (j)e-: o.* Zakład Narodowy im. Ossolińskich, Wydawn. Polskiej Akademii Nauk.

Požgaj Hadži, V., T. B. Bulc, and V. Miheljak (2013). "Srpskohrvatski jezik iz slovenske perspektive". In: *Jezik između lingvistike i politike.* Ed. by V. P. Hadži. Beograd, Biblioteka XX vek, pp. 37–65.

R Core Team (2020). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Raymond, W. D., R. Dautricourt, and E. Hume (2006). "Word-internal/t, d/deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors". In: *Language Variation and Change* 18.1, pp. 55–97. DOI: 10.1017/S0954394506060042.

Rebernik, T., J. Jacobi, R. Jonkers, A. Noiray, and M. Wieling (2021). "A review of data collection practices using electromagnetic articulography". In: *Laboratory Phonology* 12.1, pp. 1–42. DOI: 10.5334/labphon.237.

Recasens, D. and A. Espinosa (2009). "An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan". In: *The Journal of the Acoustical Society of America* 125.4, pp. 2288–2298. DOI: 10.1121/1.3089222.

Recasens, D., M. D. Pallarès, and J. Fontdevila (1997). "A model of lingual coarticulation based on articulatory constraints". In: *The Journal of the Acoustical Society of America* 102.1, pp. 544–561. DOI: 10.1121/1.419727.

Rehbein, J., J. D. ten Thije, and A. Verschik (2012). "Lingua receptiva (LaRa) - remarks on the quintessence of receptive multilingualism". In: *International Journal of Bilingualism* 16.3, pp. 248–264. DOI: 10.1177/1367006911426466.

Richmond, K., P. Hoole, and S. King (2011). "Announcing the electromagnetic articulography (day 1) subset of the mngu0 articulatory corpus". In: *Proc. Interspeech 2011*, pp. 1505–1508. DOI: 10.21437/Interspeech.2011-316.

Rouas, J.-L., J. Farinas, F. Pellegrino, and R. André-Obrecht (2005). "Rhythmic unit extraction and modelling for automatic language identification". In: *Speech Communication* 47.4, pp. 436–456. DOI: 10.1016/j.specom.2005.04.012.

Rubio, E. G. (2020). "Spanish phraseology in formal and informal spontaneous oral language production". In: *Yearbook of Phraseology* 11.1, pp. 81–106. DOI: 10.1515/phras-2020-0006.

Samardžija, M. (2015). "Gradišćanski Hrvati i gradišćanskohrvatski". In: *Poznańskie Studia Slawistyczne* 8, pp. 167–182. DOI: 10.14746/pss.2015.8.11.

Schultz, T. and A. Waibel (1998). "Multilingual and crosslingual speech recognition". In: *Proc. DARPA Workshop on Broadcast News Transcription and Understanding*, pp. 259–262. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.2484&rep=rep1&type=pdf.

Schüppert, A., N. H. Hilton, and C. Gooskens (2015). "Swedish is beautiful, Danish is ugly? Investigating the link between language attitudes and spoken word recognition". In: *Linguistics* 53.2, pp. 375–403. DOI: 10.1515/ling-2015-0003.

Schweigert, W. A. (1986). "The comprehension of familiar and less familiar idioms". In: *Journal of Psycholinguistic Research* 15.1, pp. 33–45. DOI: 10.1007/BF01067390.

Scobbie, J. M., A. Turk, C. Geng, S. King, R. Lickley, and K. Richmond (2013). "The Edinburgh Speech Production Facility DoubleTalk Corpus". In: *Proc. Interspeech 2013*. International Speech Communication Association. URL: https://www.isca-speech.org/archive/interspeech_2013/scobbie13_interspeech.html.

Sekicki, M. and M. Staudte (2018). “Eye’ll help you out! How the gaze cue reduces the cognitive load required for reference processing”. In: *Cognitive Science* 42.8, pp. 2418–2458. DOI: 10.1111/cogs.12682.

Sereno, J. and A. Jongman (1992). “Phonetic priming effects in auditory word recognition”. In: *Working Papers of the Cornell Phonetics Laboratory* 7, pp. 151–177.

Seyfarth, S. (2014). “Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation”. In: *Cognition* 133.1, pp. 140–155. DOI: 10.1016/j.cognition.2014.06.013.

Shannon, C. E. (1948). “A mathematical theory of communication”. In: *The Bell System Technical Journal* 27.3, pp. 379–423. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

Shannon, C. E. and W. Weaver (1949). *The mathematical theory of communication.* Urbana, University of Illinois Press.

Shaw, J. A., C. Han, and Y. Ma (2014). “Surviving truncation: Informativity at the interface of morphology and phonology”. In: *Morphology* 24.4, pp. 407–432. DOI: 10.1007/s11525-014-9249-5.

Shaw, J. A. and S. Kawahara (2019). “Effects of surprisal and entropy on vowel duration in Japanese”. In: *Language and Speech* 62.1, pp. 80–114. DOI: 10.1177/0023830917737331.

Skirgård, H., S. G. Roberts, and L. Yencken (2017). “Why are some languages confused for others? Investigating data from the Great Language Game”. In: *PloS One* 12.4, pp. 1–35. DOI: 10.1371/journal.pone.0165934.

Sławski, F. (1962). *Zarys dialektologii południowosłowiańskiej z wyborem tekstów gwarowych.* Warszawa, PAN.

Sławski, F. (1988). “Języki słowiańskie”. In: *Języki indoeuropejskie.* Ed. by L. Bednarczuk. Vol. 2. Warszawa, PWN, pp. 907–1005.

Sloboda, M. and M. Nábělková (2013). “Receptive multilingualism in ‘monolingual’ media: managing the presence of Slovak on Czech websites”. In: *International Journal of Multilingualism* 10.2, pp. 196–213. DOI: 10.1080/14790718.2013.789523.

Slowiaczek, L. M. and D. B. Pisoni (1986). “Effects of phonological similarity on priming in auditory lexical decision”. In: *Memory & Cognition* 14.3, pp. 230–237. DOI: 10.3758/BF03197698.

Stankiewicz, E. (1986). *The Slavic Languages: Unity in Diversity.* Walter de Gruyter.

Stenger, I. (2019). *Zur Rolle der Orthographie in der slavischen Interkomprehension mit besonderem Fokus auf die kyrillische Schrift.*

Saarbrücken, Universaar. DOI: http://dx.doi.org/10.22028/D291-32355.

Stenger, I. and T. Avgustinova (2021). "On Slavic cognate recognition in context". In: *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference 'Dialogue'*. Ed. by V. P. Selegej. Vol. 20, pp. 660–668. DOI: 10.28995/2075-7182-2021-20-660-668.

Stenger, I., T. Avgustinova, and R. Marti (2017). "Levenshtein distance and word adaptation surprisal as methods of measuring mutual intelligibility in reading comprehension of Slavic languages". In: *Computational Linguistics and Intellectual Technologies: International Conference 'Dialogue' Proceedings*. Ed. by V. P. Selegej. Vol. 1, pp. 304–317. URL: https://www.dialog-21.ru/media/3953/stengerietal.pdf.

Stevens, K. N. (1989). "On the quantal nature of speech". In: *Journal of Phonetics* 17.1-2, pp. 3–45. DOI: 10.1016/S0095-4470(19)31520-7.

Stevens, K. N. and A. S. House (1963). "Perturbation of Vowel Articulations by Consonantal Context: An Acoustical Study". In: *Journal of Speech and Hearing Research* 6.2, pp. 111–128. DOI: 10.1044/jshr.0602.111.

Stieber, Z. (1965). *Zarys dialektologii języków zachodnio-słowiańskich*. Warszawa, PWN.

Sussex, R. and P. Cubberley (2006). *The Slavic Languages*. Cambridge University Press.

Sussman, H. M., H. A. McCaffrey, and S. A. Matthews (1991). "An investigation of locus equations as a source of relational invariance for stop place categorization". In: *The Journal of the Acoustical Society of America* 90.3, pp. 1309–1325. DOI: 10.1121/1.401923.

Swarte, F. H. E. (2016). "Predicting the mutual intelligibility of Germanic languages from linguistic and extra-linguistic factors". Rijksuniversiteit Groningen. PhD thesis.

Swinney, D. A. and A. Cutler (1979). "The Access and Processing of Idiomatic Expressions". In: *Journal of Verbal Learning and Verbal Behavior* 18.5, pp. 523–534. DOI: 10.1016/S0022-5371(79)90284-6.

Tabain, M. (2000). "Coarticulation in CV syllables: a comparison of locus equation and EPG data". In: *Journal of Phonetics* 28.2, pp. 137–159. DOI: 10.1006/jpho.2000.0110.

Tanenhaus, M. K., M. J. Spivey-Knowlton, K. M. Eberhard, and J. C. Sedivy (1995). "Integration of Visual and Linguistic Information in

Spoken Language Comprehension". In: *Science* 268.5217, pp. 1632–1634. DOI: 10.1126/science.7777863.

Tang, C. and V. J. van Heuven (2007). "Mutual intelligibility and similarity of Chinese dialects: Predicting judgments from objective measures". In: *Linguistics in the Netherlands*. Ed. by B. Los and M. van Koppen. Vol. 24. 1. Amsterdam, John Benjamins, pp. 223–234.

Thije, J. D. ten and L. Zeevaert, eds. (2007). *Receptive multilingualism: Linguistic analyses, language policies and didactic concepts*. John Benjamins.

Thomas, E. (2010). *Sociophonetics: An introduction*. Palgrave Macmillan.

Tilsen, S. (2014). "Selection and coordination of articulatory gestures in temporally constrained production". In: *Journal of Phonetics* 44, pp. 26–46. DOI: 10.1016/j.wocn.2013.12.004.

Tomaschek, F., D. Arnold, K. Sering, B. V. Tucker, J. van Rij, and M. Ramscar (2021). "Articulatory Variability is Reduced by Repetition and Predictability". In: *Language and Speech* 64.3, pp. 654–680. DOI: 10.1177/0023830920948552.

Tomaschek, F., B. V. Tucker, M. Fasiolo, and R. H. Baayen (2018). "Practice makes perfect: the consequences of lexical proficiency for articulation". In: *Linguistics Vanguard* 4.s2. DOI: 10.1515/lingvan-2017-0018.

Tomaschek, F., M. Wieling, D. Arnold, and R. H. Baayen (2013). "Word frequency, vowel length and vowel quality in speech production: An EMA study of the importance of experience". In: *Proc. Interspeech 2013*. DOI: 10.21437/Interspeech.2013-347.

Topolińska, Z. (1980). "Kashubian". In: *The Slavic Literary Languages: Formation and Development*. Ed. by A. M. Schenker and E. Stankiewicz. New Haven, Yale Concilium on International and Area Studies, pp. 183–194.

Trautmann, R. (1923). *Baltisch-Slavisches Wörterbuch*. Göttingen, Vandenhoeck & Ruprecht.

Turk, A. (2010). "Does prosodic constituency signal relative predictability? A Smooth Signal Redundancy hypothesis". In: *Laboratory Phonology* 1.2, pp. 227–262. DOI: 10.1515/labphon.2010.012.

Urbańczyk, S. (1953). *Zarys dialektologii polskiej*. Warszawa, Państwowe Wydawnictwo Naukowe.

Vanhove, J. (2014). "Receptive multilingualism across the lifespan: Cognitive and linguistic factors in cognate guessing". PhD thesis. University of Fribourg. URL: https://doc.rero.ch/record/210293/files/VanhoveJ.pdf.

Vega-Moreno, R. E. (2001). "Representing and processing idioms". In: *UCL Wording Papers in Linguistics* 13, pp. 73–107.

Verschik, A. (2012). "Practising receptive multilingualism: Estonian–Finnish communication in Tallinn". In: *International Journal of Bilingualism* 16.3, pp. 265–286. DOI: 10.1177/1367006911426465.

Voegelin, C. F. and Z. S. Harris (1951). "Methods for determining intelligibility among dialects of natural languages". In: *Proceedings of the American Philosophical Society* 95.3, pp. 322–329.

Vuković, P. (2015). "Buniewcy z Bački – język i tożsamość". In: *Poznańskie studia slawistyczne* 8, pp. 239–256. DOI: 10.14746/pss.2015.8.16.

Walczak-Mikołajczakowa, M. (2015). "Bułgarzy banaccy–fenomen trwałości kultury". In: *Poznańskie Studia Slawistyczne* 8, pp. 229–237. DOI: 10.14746/pss.2015.8.15.

Weber, A. and A. Cutler (2004). "Lexical competition in non-native spoken-word recognition". In: *Journal of Memory and Language* 50.1, pp. 1–25. DOI: 10.1016/S0749-596X(03)00105-0.

Wilson, K. and P. Foulkes (2014). "Borders, variation, and identity: Language analysis for the determination of origin (LADO)". In: *Language, borders and identity*. Ed. by D. Watt and C. Llamas. Edinburgh University Press, pp. 218–229.

Wolff, H. (1959). "Intelligibility and inter-ethnic attitudes". In: *Anthropological Linguistics* 1, pp. 34–41.

Wrench, A. (1999). *MOCHA-TIMIT*. Articulatory database. Queen Margaret University College, Edinburgh. URL: https://www.cstr.ed.ac.uk/research/projects/artic/mocha.html.

Zissman, M. A. and K. M. Berkling (2001). "Automatic language identification". In: *Speech Communication* 35.1-2, pp. 115–124. DOI: 10.1016/S0167-6393(00)00099-6.

Intercomprehension refers to a communication practice in which speakers use closely related languages. We know that the degree of mutual intelligibility differs according to the stimulus modality. This work aims to define the linguistic features which contribute to and impede cross-lingual understanding of speech via production and perception studies involving speakers of four Slavic languages. The current study combines the methodological apparatus from acoustic phonetics and information theory to provide evidence for mutual intelligibility on various levels of language processing. It concludes that the degree of mutual understanding does not always correspond to typological divisions of tested languages. The results presented here suggest that intercomprehension is often driven by unit (un)expectedness rather than the phonetic resemblance of a perceived stimulus and its equivalence in the native lexicon of speakers.

ISBN 978-3-86223-316-8

9 783862 233168 >