

Русско-болгарская взаимопонятность в свете лингвистических и статистических моделей славянской рецептивной многоязычности

Таня Августинова, Ирина Штенгер (Universität des Saarlandes, Saarbrücken)

Povzetek

Računalniško modeliranje jezikovnih okoliščin, v katerih so slovanski jeziki medsebojno razumljivi, vključuje sistematsko fokusiranje na klasična slavistična znanja primerjalne slovanske historične slovnice in sinhronega protistavnega opisovanja jezikovnih parov. Fenomen medjezikovnega razumevanja vključuje intuitivno razumevanje: govorec jezika L1, ki ne obvlada jezika L2, se opira na jezikovni sistem jezika L1. Gre za specifičen način rabe jezikovne kompetence, ki se kaže kot receptivna oziroma pasivna večjezičnost. Stopnja medsebojne razumljivosti med tesno povezanimi jeziki, kot sta bolgarski in ruski, ustreza objektivno merljivi jezikovni razdalji med njimi. V primerjalnem historičnem razvoju skupni slovanski okvir daje možnost abstraktne formulacije pojavov, ki nas zanimajo. Sinhrona perspektiva pa nudi konkretno osnovo za raziskovanje jezikovne bližine med bolgarščino in ruščino v pravopisu/pisavi, leksiki in slovnici.

Ključne besede: medjezikovno razumevanje, slovanski jeziki, jezikovna razdalja, rusko-bolgarska medjezikovna razumljivost, modeli za receptivno večjezičnost

Abstract

Computational modelling of the observed mutual intelligibility of Slavic languages unavoidably requires systematic integration of classical Slavistics knowledge from comparative historical grammar and traditional contrastive description of language pairs. The phenomenon of intercomprehension is quite intuitive: speakers of a given language L1 understand another closely related language (variety) L2 without being able to use the latter productively, i.e. for speaking or writing. This specific mode of using the human linguistic competence manifests itself as receptive multilingualism. The degree of mutual understanding of genetically closely related languages, such as Bulgarian and Russian, corresponds to objectively measurable distances at different linguistic levels. The common Slavic basis and the comparative-synchronous perspective allow us to reveal Bulgarian-Russian linguistic affinity with regard to spelling, vocabulary and grammar.

Key words: intercomprehension, Slavic languages, linguistic distance, Russian-Bulgarian mutual intelligibility, models of receptive multilingualism

1 Введение

Благодаря схожести лексики и грамматики генетически близких языков происходит следующее: человек, воспринимающий неродной, хотя и родственный язык, вольно или невольно использует знание своего родного языка при разборе и интерпретации воспринимаемого им другого языка. При таком режиме неизбежно порождается определённая информационная избыточность или же возникает информационная недостаточность. Например, каким образом носитель русского языка, с его богатой именной морфологией, поймёт текст на болгарском языке, где склонение почти исчезло? И напротив, как русский воспримет разветвлённую систему болгарских времён при том, что в его родном языке он обходится одной формой для обозначения прошедшего времени?

Межъязыковое понимание (*intercomprehension* Doyé 2005), рецептивная многоязычность (*rezeptive Mehrsprachigkeit* Braunmüller & Zeevaert 2001) или семи-коммуникация (*semi-communication* Haugen 1966) представляют собой особый вид общения между носителями схожих языков, при котором каждый из участников коммуникации использует свой язык, пытаясь понять или извлечь информацию на незнакомом ему языке. Степень взаимопонятности (*mutual intelligibility*) между (близко)родственными языками зависит как от лингвистических, так и от экстралингвистических факторов (Gooskens 2013). К лингвистическим факторам относятся объективные сходства языков на материальном и структурном уровнях фонетики/фонологии, графики/орфографии, лексики, морфологии, синтаксиса. Экстралингвистические факторы представляют собой, с одной стороны, социальные параметры носителей языка (например, возраст, образование, языковой опыт, языковой контакт), с другой стороны когнитивные характеристики (например, интеллектуальные способности носителя языка). Предполагается, что чем больше лингвистического сходства между двумя языками, тем выше будет степень их взаимопонятности. При этом взаимопонятность между языками может быть асимметричной, когда носители языка А понимают язык Б лучше, чем носители языка Б понимают язык А (Ringbom 2007). В последние годы в экспериментальных исследованиях в разной степени уделялось внимание межъязыковому пониманию между скандинавскими языками (Gooskens 2006, Kürschner et al. 2008), между германскими языками (Gooskens & Swarte 2017, Möller & Zeevaert 2015), между славянскими языками (Golubović & Gooskens 2015, Reichert 2013).

Рецептивная, или пассивная, многоязычность проявляется в ситуации, когда читателю или слушателю необходимо извлечь информацию из текста на

неродном языке без реального владения последним. На наш взгляд, степень взаимопонятности в различных парах славянских языков усиливается наличием общего словаря и схожих грамматических структур, а также осознанием факта этого наличия. Когда носитель языка L1 получает сообщение на незнакомом ему языке Lx, это сообщение обогащается наличным знанием об L1 и порождает гипотезы о форме и структуре сообщения на Lx, в которых, оптимально, вовлечено осознание процессов интерференции между L1 и Lx. При изучении употребления языка в режиме интерференционного понимания удаётся обнаружить межъязыковую “толерантность” к незнакомой кодировке, а также, в зависимости от конкретной языковой пары, асимметричность степени понятности. Коль скоро полноценная интерпретация формы и структуры невозможна, разумной стратегией интерференционной коммуникации становится фрагментный анализ, который сосредоточивается на понятных деталях текста и игнорирует непонятные.

В проекте INCOMSLAV (Avgustinova et al. 2014–2018) проводился анализ лингвистических факторов с использованием вероятностно-статистических методов теории информации (Shannon 1948), а также методов математического моделирования (Levenshtein 1966). Результаты статистических анализов свидетельствуют о том, что, например, болгарский и русский языки являются более близкими на уровне орфографии, чем польский и чешский (Fischer et al. 2015, Stenger et al. 2017b). В области лексики расхождения между парами языков уже не такие значительные (Jágrová et al. 2017). Для проведения экспериментов с носителями славянских языков с целью сбора материала по пониманию незнакомых, но (близко)родственных языков при чтении была создана веб-платформа (<http://intercomprehension.coli.uni-saarland.de>). В данной статье мы представляем результаты проведённых экспериментов в режиме онлайн с носителями болгарского и русского языков, а также два метода оценки степени языковой близости – расстояние Левенштейна и меру неожиданности адаптации слова – в качестве потенциальных параметров определения успешности межславянского взаимопонимания.

2 Экспериментальное исследование

На базе веб-платформы¹ были проведены эксперименты по свободному переводу с болгарского языка на русский и с русского языка на болгарский. В качестве информантов выступили 71 носитель болгарского языка и 94

1 На данный момент экспериментальная веб-платформа переведена на 11 славянских языков. Участие в экспериментах приняли около 1600 носителей славянских языков (03.12.2018).

носителя русского языка. С целью исключения эффекта обучения были отобраны те участники, которые выполнили данный эксперимент как первый, а также те носители языка, которые указали, что не владеют языком-стимулом. В итоге анализ собранного материала базируется на ответах 37 носителей болгарского языка (31 женщина и 6 мужчин в возрасте от 18 до 41 года, средний возраст 27 лет) и 40 носителей русского языка (32 женщины и 8 мужчин в возрасте от 18 до 71 года, средний возраст 33 года).

Эксперименты проводились в качестве свободного перевода когнатов, то есть однокоренных слов общего происхождения, с языка-стимула на родной язык. При предъявлении когнатов в письменной форме рассматривалась в первую очередь орфографическая понятность стимулов вне контекста. Основная гипотеза проведения данного типа экспериментов заключается в том, что успешное узнавание определённого количества когнатов – это ключ к пониманию текста на незнакомом, но (близко)родственном языке.

Материалом исследования служили 120 пар когнатов. Основанием для сравнения были болгарско-русские диахронические орфографические соответствия, типа *ла:оло*, например, в *хлад–холод*. Выбор 120 пар когнатов был выполнен на базе следующих списков слов: интернационализмы и слова общеславянской лексики проекта EuroComSlav², а также список Сводеша³ (подробнее о материале см. в Fischer et al. 2015, Stenger et al. 2017a). 120 когнатов⁴ предъявлялись в качестве стимулов по 60 слов в двух экспериментах. Стимулы показывались на экране персонального компьютера поочерёдно в письменной форме. На каждый стимул участникам эксперимента предоставлялось 10 секунд для письменного ответа. По истечении 10 секунд на экране появлялся следующий стимул. Реакции респондентов автоматически распознавались как правильные или неправильные ответы, затем происходил ручной контроль неправильных ответов на предмет опечаток. Для автоматического распознавания ответов были составлены дополнительные списки возможных альтернативных ответов. Так, например, болгарское слово *път* может быть переведено на русский язык как *путь* и как *дорога*. Оба ответа считались правильными. На основе полученных данных формировалась база правильных и неправильных ответов, а также определялось их количество. В Таблице 1 представлен средний показатель правильных ответов 37-ми носителей болгарского языка, при общем

2 <http://www.eurocomslav.de/kurs/pwslav.htm>;
<http://www.eurocomslav.de/kurs/iwslav.htm> (11.07.2015).

3 http://en.wiktionary.org/wiki/Appendix:Swadesh_lists_for_Slavic_languages.

4 118 существительных и два количественных числительных.

количестве экспериментальных реакций 4400, и 40-а носителей русского языка, при общем количестве экспериментальных реакций 4800.

Таблица 1: Результаты экспериментов по свободному переводу

родной язык	язык-стимул	
	русский	болгарский
русский		71,33%
болгарский	74,67%	

Как показывают результаты экспериментов по свободному переводу, орфографическая взаимопонятность изолированных 120 когнатов достаточно высока, что вполне объяснимо, так как болгарский и русский языки отличаются большим сходством на уровне графики и орфографии. В отличие от русского алфавита, где 33 буквы, болгарский состоит из 30 букв – здесь нет букв *э, ы, ё*. Однако буква *ь* (“ер голям”) обозначает в болгарском языке характерный гласный, а не является твёрдым знаком, как в русском языке. В целом соотношение между буквами алфавита и звуками речи такое же, что и в русском языке, хотя есть и некоторые особенности (Иванова et al. 2011, Маслов 1981).

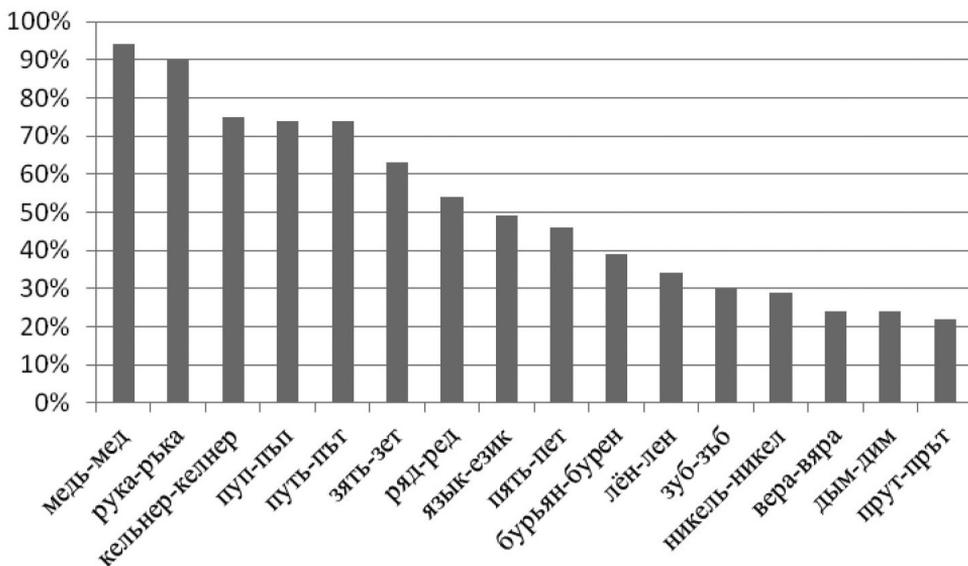
Следует отметить, что носители болгарского языка успешно справились с переводом русских стимулов в среднем на 3,34% лучше, чем носители русского языка с переводом болгарских стимулов. Однако данная разница оказалась статистически незначимой ($t = -0,95$, $df = 238$, $p = 0,341$).⁵

С целью детального изучения асимметрии взаимопонятности мы провели сравнительный анализ правильных ответов между болгарскими и русскими стимулами, чтобы определить, какие стимулы вызывают наибольшие затруднения у носителей соответствующего языка, а также какие именно диахронические орфографические соответствия наиболее непонятны при чтении незнакомого языка. Выяснилось, что 70 когнатов из 120 были успешнее переведены носителями болгарского языка, чем носителями русского языка, а 47 когнатов были лучше поняты носителями русского языка, чем болгарского. Три пары когнатов были практически одинаково хорошо поняты как носителями болгарского языка (94,74%), так и носителями русского языка (95,00%): *морков–морковь, музыка–музыка, яйце–яйцо*.

5 t-критерий используется для определения достоверности разницы средних значений в двух выборках, df – число степеней свободы, т.е. количество значений в итоговом вычислении статистики, способных варьироваться, p – критерий для определения уровня значимости, в данном случае $p > 0,05$.

На Рисунке 1 представлены 16 когнатов, которые были лучше переведены носителями болгарского языка. На вертикальной оси указана разница (в процентах) правильных переводов; в данную группу включены ответы, разница которых выше 22%. Девять русско-болгарских орфографических соответствий не представляли затруднений для носителей болгарского языка при понимании русских стимулов: *дь:д* (1)⁶, *е:я* (1), *ё:е* (1), *ль:л* (2), *рь:р* (1), *ть:т* (3), *у:ь* (5), *ы:и* (2), *я:е* (5). Носители болгарского языка испытывали в основном затруднения с заменой гласных букв, например, русских гласных букв *е, о, у* на болгарскую гласную букву *ь*, а также некоторые сложности с полногласием русского языка в сочетаниях *-оло-*, *-оро-*.

Рисунок 1: Когнаты, которые были переведены лучше носителями болгарского языка

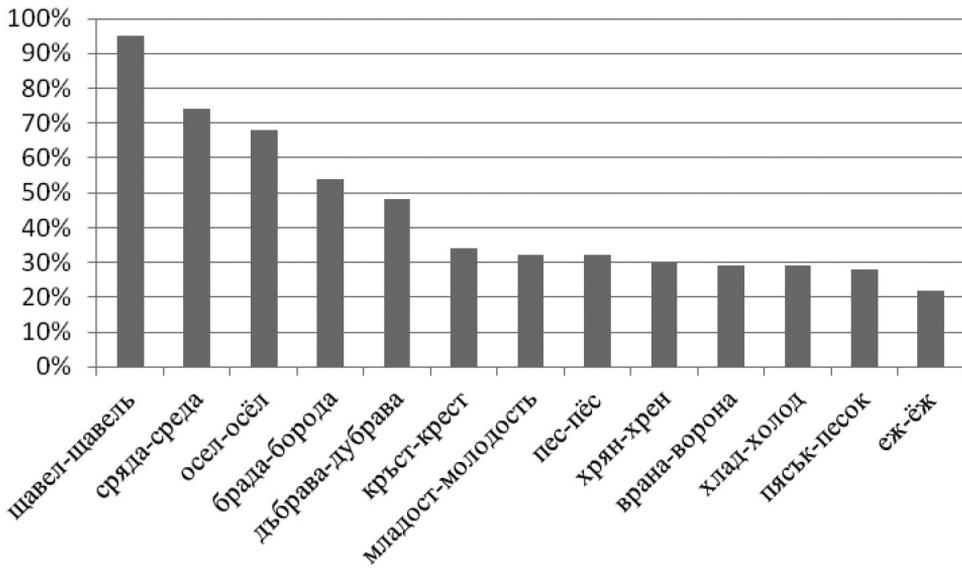


На Рисунке 2 представлены 13 когнатов, которые были переведены лучше носителями русского языка. На вертикальной оси указана разница (в процентах) правильных переводов; в данную группу вошли ответы, разница которых выше 22%. Восемь болгарско-русских орфографических соответствий не представляли трудностей для носителей русского языка при понимании болгарских стимулов: *е:ё* (3), *л:ль* (1), *ла:оло* (2), *ра:оро* (2), *ь:е* (1),

6 В круглых скобках указана частотность орфографических соответствий в группе когнатов.

ь:о (1), ь:у (1), я:е (3). Носители русского языка испытывали трудности при замене гласных букв, например, болгарской гласной буквы ь на русскую гласную букву у, а также с характерной особенностью болгарского языка, что на конце слова и перед другим согласным звуком могут быть только твёрдые согласные, например, *път* – *путь*.

Рисунок 2: Когнаты, которые были переведены лучше носителями русского языка



Для выше представленных примеров характерна разница правильных ответов выше 22%, например, русский стимул *прут* был переведён носителями болгарского языка в среднем на 22% успешнее, чем соответствующий болгарский стимул *прът* был понят носителями русского языка. Самая большая разница правильных ответов характерна для пары *щавел* – *щавель*. В этом случае носители русского языка на 95% лучше справились с заданием по свободному переводу, чем носители болгарского языка, что можно объяснить наличием в болгарском языке более частотного неоднокоренного синонима “киселец”.

Как было сказано выше, носители болгарского языка справились с заданием по свободному переводу в среднем лишь с небольшим преимуществом (на 3,34% лучше) по сравнению с носителями русского языка. Детальный анализ правильных ответов показывает, что в определённых случаях наблюдается асимметрия взаимопонятности от 22% до 95% как со стороны русского

языка, так и со стороны болгарского языка. Однако количество когнатов в данных группах незначительно: 16 русских стимулов и 13 болгарских стимулов из 120 пар когнатов.

3 Методы оценки близости языков

3.1 Расстояние Левенштейна

Расстояние Левенштейна (Levenshtein 1966) – это минимальное количество операций вставки одного символа, удаления одного символа или замены одного символа на другой, необходимых для превращения одной строки в другую. Алгоритм Левенштейна может быть модифицирован с учётом лингвистических особенностей, например, согласные буквы могут заменяться только согласными, а гласные только гласными. При этом цена каждой операции, т.е. вставки, удаления или замены неидентичных символов составляет, как правило, 1, например, замена гласной буквы *a* на *o*⁷, а идентичных символов 0. В Таблице 2 представлен пример расчёта расстояния Левенштейна для болгарско-русской пары когнатов *младост–молодость*.

Таблица 2: Расстояние Левенштейна (*Levenshtein distance LD*) для болгарско-русской пары когнатов *младост–молодость*

	1	2	3	4	5	6	7	8	9	LD
болгарский	м		л	а	д	о	с	т		
русский	м	о	л	о	д	о	с	т	ь	
	0	1	0	1	0	0	0	0	1	3

Расстояние Левенштейна (LD) для пары когнатов *младост–молодость* равно 3. Далее расстояние нормализуется по длине строки (9): $3/9=0,33$ или 33%. Полученные 33% рассматриваются как орфографическая дистанция и выступают в качестве статистической оценки сложности понимания русского слова *молодость* носителем болгарского языка или понимания болгарского слова *младост* носителем русского языка.

С помощью алгоритма Левенштейна были рассчитаны орфографические дистанции для 120 пар когнатов и соотнесены с результатами экспериментов.

7 Замена русской буквы *ё* на болгарскую букву *e* алгоритмом оценивалась как 0,5.

Метрический анализ Левенштейна, часто используемый для выявления предсказуемости фонетического и орфографического сходства между когнатами, основан на измерении математического расстояния и, поэтому, даёт абсолютно симметричные результаты. Однако, как было отмечено выше, при рецептивной многоязычности ожидается асимметрия взаимопонятности в языковой паре. Результаты проведённых экспериментов с носителями болгарского и русского языков показали, что асимметричность орфографической взаимопонятности стимулов настолько минимальна, что статистически оказалась незначимой. Однако орфографическую асимметрию между родственными языками можно выявить с помощью следующего метода статистической оценки близости языков.

3.2 Мера неожиданности адаптации слова

Дополнительный параметр, который применяется нами для оценки близости между родственными языками, – это мера неожиданности адаптации слова (*word adaptation surprisal*), основанная на теории передачи информации Шеннона (Shannon 1948). Мера неожиданности адаптации слова позволяет статистически рассчитать орфографическую асимметрию, а также оценить затруднения, испытываемые в ходе соотнесения одной орфографической системы с другой (Stenger et al. 2017b). Исходная гипотеза заключается в том, что чем больше мера неожиданности адаптации слова, тем сложнее понять когнат незнакомого, но (близко)родственного языка. Мера неожиданности адаптации слова складывается из мер неожиданности адаптации графических символов между когнатами (которые рассчитываются с помощью двоичного логарифма – см. формулу в Stenger et al. 2017a, Stenger et al. 2017b).

В Таблице 3 представлены меры неожиданности адаптации графических символов в болгарско-русской паре когнатов *младост–молодость* (расчёт мер неожиданности в битах выполнен на базе 120 пар когнатов для двух языков-стимулов). Согласно приведённым данным мера неожиданности болгарской адаптации русского стимула *молодость* составляет 5,9 бит, а мера неожиданности русской адаптации болгарского стимула *младост* – 4,41 бит. Далее, как и в случае с орфографической дистанцией, мера неожиданности адаптации слова нормализуется по длине строки: $5,9/9 = 0,66$ бит для адаптации стимула *молодость*, а $4,41/9 = 0,49$ бит для *младост*. Следовательно, носители русского языка должны испытывать меньше затруднений при понимании болгарского стимула, чем носители болгарского языка при переводе русского стимула. Действительно, носители русского языка в среднем на 32% были успешнее при понимании болгарского стимула

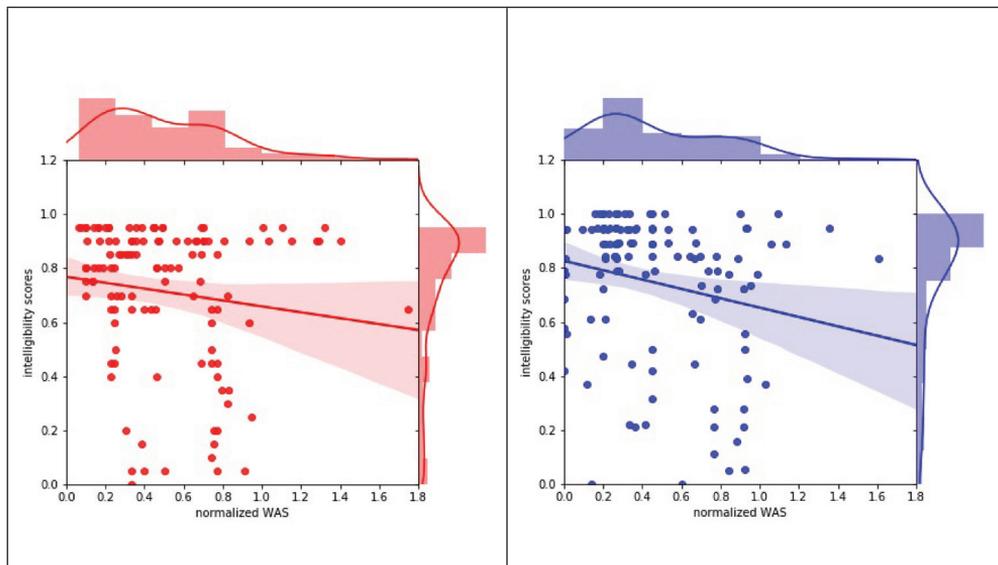
младост, чем носители болгарского языка при переводе русского стимула *молодость*.

Таблица 3: Мера неожиданности адаптации графических символов в болгарско-русской паре когнатов *младост* – *молодость*

позиция в строке	1	2	3	4	5	6	7	8	9
болгарский	м		л	а	д	о	с	т	
русский	м	о	л	о	д	о	с	т	ь
язык-стимул: русский	0	2,34	0,03	2,24	0	1,29	0	0	0
язык-стимул: болгарский	0	2,05	0	1,82	0	0	0	0	0,54

Меры неожиданности адаптации слова были рассчитаны для каждой пары 120 когнатов в зависимости от языка-стимула и соотнесены с результатами экспериментов. Приведённые в Рисунке 4 данные позволяют утверждать, что значимая отрицательная корреляция между мерой неожиданности адаптации слова и пониманием стимула незнакомого, но (близко) родственного языка была обнаружена только в группе с русским языком-стимулом (4б) ($r = -0,22$, $R^2 = 0,05$, $p < 0,05$). В группе с болгарским языком-стимулом (4а) отрицательная корреляция между мерой неожиданности адаптации слова и пониманием стимула оказалось незначимой ($r = -0,13$, $R^2 = 0,02$, $p = 0,14$). Таким образом, можно предварительно сказать, что такой параметр как мера неожиданности адаптации слова оказался значимым только для носителей болгарского языка, что объясняет понимание изолированных когнатов при чтении русского языка, но только на 5% (см. $R^2 = 0,05$). Представленный метод оценки близости языков позволяет также оценить орфографическую асимметрию между болгарским и русским языками. Средний показатель меры неожиданности адаптации слова для носителей болгарского языка составляет 0,46 бит, а для носителей русского языка 0,50 бит. Согласно статистическим данным носители болгарского языка должны иметь минимальное преимущество при чтении русских стимулов по отношению к носителям русского языка при понимании болгарских стимулов. Таким образом, статистический расчёт среднего показателя меры неожиданности слова подтверждает незначительную асимметрию взаимопонятности между болгарским и русским языками на материале проведённых экспериментов (см. раздел 2).

Рисунок 4: Мера неожиданности адаптации слова
(normalized W[ord]A[daptation]S[urprisal])
как метод оценки орфографической понятности (intelligibility score)
4а: болгарский – язык-стимул 4б: русский – язык-стимул



4 Заключение

Проведённые эксперименты с носителями болгарского и русского языков показали интересные результаты, связанные с русско-болгарской взаимопонятностью в свете рецептивной многоязычности. Хотя болгарский язык относится к южнославянской группе, а русский к восточнославянской, результаты проведённых нами экспериментов свидетельствуют о том, что для данных языков характерен достаточно высокий уровень орфографической взаимопонятности. Носители болгарского языка справились с заданием по свободному переводу изолированных когнатов на 3,34% лучше, чем носители русского языка. Минимальная разница в правильных ответах оказалась статистически незначимой, что позволяет говорить об отсутствии асимметрии между данными языками на орфографическом уровне. Следует, однако, отметить, что детальный анализ правильных ответов показал, что ряд когнатов понимается носителями болгарского языка лучше, чем носителями русского языка, и наоборот, о чём свидетельствует разница правильных ответов от 22% до 95% в зависимости от языка-стимула, хотя количество стимулов в данных группах незначительно: 16 русских стимулов и 13 болгарских стимулов из 120 пар когнатов.

Далее нами были представлены два метода оценки степени близости незнакомых языков – расстояние Левенштейна и мера неожиданности адаптации слова – как потенциальные параметры определения успеха межславянского взаимопонимания. Орфографическая дистанция, рассчитанная с помощью алгоритма Левенштейна, оказалась статистически значимым параметром, объясняющим понимание при чтении изолированных когнатов. Это означает, что чем больше орфографическая дистанция между когнатами, тем сложнее их понять. Что касается меры неожиданности адаптации слова, то данный параметр оказался статистически значимым только для понимания русских стимулов носителями болгарского языка. В случае понимания болгарских стимулов носителями русского языка данный параметр оказался статистически незначимым. На основе этого можно сделать вывод, что такие факторы как регулярность и частотность русско-болгарских орфографических соответствий, которые лежат в основе подсчёта меры неожиданности адаптации слова, в определённой степени оказывают влияние на понимание русских стимулов, хотя этот эффект весьма слабый (5%).

С помощью среднего показателя меры неожиданности адаптации слова мы определили незначительную асимметрию между болгарским и русским языками на орфографическом уровне. Данный статистический анализ подтверждается экспериментальными данными, которые показывают незначительное преимущество носителей болгарского языка в понимании русского перед носителями русского языка в понимании болгарского.

В заключение заметим, что понимание изолированных когнатов незнакомо-го, но (близко)родственного языка зависит и от ряда других факторов, например, частотности когнатов родного языка, количества орфографических соседей⁹ в родном языке по отношению к стимулам незнакомо-го языка, а также от самих орфографических соответствий, их природы, места в слове (в начале, в середине или в конце), их количества в слове. Как показывают исследования, большее или меньшее влияние одного или другого фактора зависит как от анализируемой пары когнатов в частности, так и от анализируемой пары языков в целом.

9 Орфографическими соседями называются слова, отличающиеся друг от друга на одну букву, например *крот – кот, крот – кров*.

Литература

- Иванова, Елена Ю., Зоя К. Шанова & Димитринка Димитрова, 2011: *Болгарский язык*. Санкт-Петербург: Издательство Каро.
- Маслов, Юрий С., 1981: *Грамматика болгарского языка*. Москва: «Высшая школа».
- Avgustinova, Tania, Dietrich Klakow & Roland Marti, 2014–2018: *Mutual Intelligibility and Surprisal in Slavic Intercomprehension*. Project INCOMSLAV, SFB 1102 Information Density and Linguistic Encoding, Collaborative Research Center at Saarland University: funded by the German Science Foundation (DFG).
- Braunmüller, Kurt & Ludger Zeevaert L., 2001: *Semikommunikation, rezep tive Mehrsprachigkeit und verwandte Phänomene. Eine bibliographische Bestandsaufnahme* (Arbeiten zur Mehrsprachigkeit, Folge B, 19). Hamburg: Universität Hamburg.
- Doyé, Peter, 2005: Intercomprehension. *Guide for the Development of Language Education Policies in Europe: From Linguistic Diversity to Plurilingual Education* (Reference Studies). Strasbourg: Council of Europe, 105–123.
- Fischer, Andrea, Klára Jágrová, Irina Stenger, Tania Avgustinova, Dietrich Klakow & Roland Marti, 2015: An orthography transformation experiment with Czech–Polish and Bulgarian–Russian parallel word sets. Bernadette Sharp, Wiesław Lubaszewski & Rodolfo Delmonte (eds.), *Natural Language Processing and Cognitive Science 2015 Proceedings*. Venezia: Libreria Editrice Cafoscarina, 115–126.
- Golubović, Jelena & Charlotte Gooskens, 2015: Mutual intelligibility between West and South Slavic languages. In: *Russian Linguistics* 39. Springer DOI 10.1007/s11185-015-9150-9, 351–373.
- Gooskens, Charlotte, 2006: Linguistic and extra-linguistic predictors of Inter-Scandinavian intelligibility. van de Weijer, Jeroen & Bettelou Los (eds.), *Linguistics in the Netherlands* 23. Amsterdam: John Benjamins, 101–113.
- Gooskens, Charlotte, 2013: Experimental methods for measuring intelligibility of closely related language varieties. Bayley, Robert, Richard Cameron & Ceil Lucas (eds.), *Handbook of Sociolinguistics*. Oxford: Oxford University Press, 195–213.
- Gooskens, Charlotte & Femke Swarte, 2017: Linguistic and extra-linguistic predictors of mutual intelligibility between Germanic languages. *Nordic Journal of Linguistics* 40(2), 123–147.

- Haugen, Einar, 1966: Semicommunication: The language gap in Scandinavia. *Sociological Inquiry* 36, 280–297.
- Jágrová, Klára, Irina Stenger, Roland Marti & Tania Avgustinova, 2017: Lexical and orthographic distances between Bulgarian, Czech, Polish, and Russian: A comparative analysis of the most frequent nouns. Joseph Emonds & Markéta Janebová (eds.), *Language Use and Linguistic Structure. Proceedings of the Olomouc Linguistics Colloquium 2016*. Olomouc: Palacký University, 401–416.
- Kürschner, Sebastian, Charlotte Gooskens & Renée van Bezooijen, 2008: Linguistic determinants of the intelligibility of Swedish words among Danes. *International Journal of Humanities and Arts Computing* 2(1/2), 83–100.
- Levenshtein, Vladimir I., 1966: Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady* 10(8), 707–710.
- Möller, Robert & Ludger Zeevaert, 2015: Investigating word recognition in intercomprehension: Methods and findings. *Linguistics* 53(2), 313–352.
- Reichert, Anna, 2013: Auditive intercomprehension between Russian and Bulgarian – degree and factors for the listening comprehension of native speakers. Juliane Besters-Dilger & Uliana Schöller. *Slavischer Sprachkontakt. Beiträge zum gleichnamigen Workshop für Studierende und Promovierende, Freiburg, 19.–20. April 2013*, 119–130. <https://freidok.uni-freiburg.de/dnb/download/9232> (Stand 28.11.2018).
- Ringbom, Håkan, 2007: *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters Ltd.
- Shannon, Claude E., 1948: A mathematical theory of communication. *Bell System Technical Journal* 27(379–423), 623–656.
- Stenger, Irina, Tania Avgustinova & Roland Marti, 2017a: Levenshtein distance and word adaptation surprisal as methods of measuring mutual intelligibility in reading comprehension of Slavic languages. Vladimir P. Selegej, (ed.) *Computational Linguistics and Intellectual Technologies: International Conference 'Dialogue 2017' Proceedings*. Issue 16(23), vol. 1, 304–317.
- Stenger, Irina, Klára Jágrová, Andrea Fischer, Tania Avgustinova, Dietrich Klakow & Roland Marti, 2017b: Modeling the Impact of Orthographic Coding on Czech-Polish and Bulgarian-Russian Reading Intercomprehension. *Nordic Journal of Linguistics* 40(2), 175–199.