**Взаимодействие лингвистических и социокогнитивных параметров при рецептивном мультилингвизме**

**Штенгер И.** (ira.stenger@mx.uni-saarland.de)1,   
**Августинова Т.** (avgustinova@coli.uni-saarland.de)1,

**Белоусов К. И.** (belousovki@gmail.com)2

**Баранов Д. А.** (baranov@semograph.com)2

**Ерофеева Е. В.** (elenerofee@gmail.com)2

1Университет земли Саар, Саарбрюккен, Германия;

2Пермский государственный национальный исследовательский университет

Receptive multilingualism or intercomprehension is a special mode of language use that characteristically involves multi-source compensatory guessing strategies. Objectively, mutual intelligibility of unknown but genetically or typologically related languages correlates with measurable similarities in lexis, phonetics, orthography and morphosyntax. Subjectively, intercomprehension is influenced by extra-linguistic factors such as experience, age, attitude, exposure. In this paper we present the results of web-based experiments obtained from Russian native speakers trying to understand written Bulgarian words out of context in a cognate guessing challenge. In particular, we investigate the correlation of linguistic intelligibility based on orthographic correspondences and string similarity with socio-cognitive factors of sex and reaction time in order to pursue a deeper understanding of the processes involved in receptive multilingualism. To organize collaborative work, an information management system is used for visualisation of the results.

**Keywords:** receptive multilingualism, cognate guessing, orthographic correspondences, Russian and Bulgarian, distance, reaction time, visualisation

**1. Введение**

Рецептивный мультилингвизм (*rezeptive Mehrsprachigkeit* Braunmüller, Zeevaert 2001), межъязыковое понимание (*intercomprehension* Doyé 2005), или семи-коммуникация (*semi-communication* Haugen 1966), представляет собой особый вид общения между носителями схожих языков, при котором каждый из участников коммуникации использует свой язык при извлечении информации, доступной 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. Предполагается, что чем сильнее выражено лингвистическое сходство между двумя языками, тем выше степень их взаимопонятности. Экстралингвистические факторы представляют собой, с одной стороны, социальные параметры носителей языка (возраст, пол, образование и др.), с другой – когнитивные характеристики, выражающиеся прежде всего в точности и быстроте реакции. Кроме того, процессы рецептивного мультилингвизма зависят от индивидуального языкового репертуара, т.е. количества известных человеку языков и степени владения каждым из них. В экспериментальных исследованиях последних лет особое внимание уделялось межъязыковому пониманию между скандинавскими языками (Gooskens 2006; Kürschner, Gooskens, van Bezooijen 2008), германскими языками (Gooskens, Swarte 2017; Möller, Zeevaert 2015), славянскими языками (Golubović, Gooskens 2015; Reichert 2013; Stenger, Avgustinova, Marti 2017).

В данной статье представлены результаты экспериментов по свободному переводу болгарских слов носителями русского языка. Исследование проведено на платформе INCOMSLAV (http://intercomprehension.coli.uni-saarland.de, Университет земли Саар, Германия), используемой для проведения онлайн-экспериментов и формирования корпуса экспериментальных реакций.  Обработка и анализ материала для настоящего исследования осуществлялись с помощью информационной системы «Семограф» (https://v2.semograph.org, Пермский государственный национальный исследовательский университет). В частности анализируется взаимосвязь таких социокогнитивных характеристик информантов, как пол и время реакции, с понятностью болгарских стимулов на базе анализа орфографических соответствий и орфографической дистанции.

**2. Экспериментальное исследование**

Предполагается, что успешное распознавание достаточно большого количества когнатов (однокоренных слов общего происхождения) – это ключ к пониманию текста на незнакомом, но (близко)родственном языке (Gooskens 2013; Kürschner, Gooskens, van Bezooijen 2008). На базе веб-платформы[[1]](#footnote-1) были проведены эксперименты по свободному переводу  когнатов с болгарского языка (языка-стимула) на русский (родной) язык.

Материалом исследования служила подборка 120 пар когнатов, включающих болгарско-русские диахронически обусловленные орфографические соответствия, типа *ла*:*оло*, например, *хлад* – *холод*. Выбор когнатов осуществлялся из списков слов-интернационализмов и слов общеславянской лексики проекта EuroComSlav, а также списка Сводеша. Все три списка были скорректированы и незначительно изменены следующим образом: если пары слов не являлись когнатами, то они удалялись из списка или по возможности дополнялись альтернативными когнатами. Например, болгарско-русская пара слов *ние* – *мы* была удалена, а болгарское слово *звяр* (вместо *животно*) было добавлено к русскому когнату *зверь* (подробнее о материале см. в Fischer et al. 2015; Stenger et al. 2017).

Отобранные когнаты (118 существительных и два количественных числительных) в письменной форме поочередно предъявлялись информантам на экране компьютера. На каждый стимул участникам эксперимента предоставлялось 10 секунд для письменного ответа, по истечении которых на экране появлялся следующий стимул. Таким образом было проведено две серии данного эксперимента (по 60 несовпадающих слов в каждой серии (см. приложение 1 и 2), порядок предъявления стимулов внутри одной серии отличался для разных респондентов).

В качестве информантов выступили 123 носителя русского языка (первая серия – 50 жен. и 14 муж., средний возраст – 31 год; вторая серия – 39 жен. и 20 муж., средний возраст – 32 года); 20 информантов (13 жен. и 7 муж.) приняли участие в обеих сериях. Перед выполнением эксперимента информанты заполняли анкету, где указывали возраст, пол, место жительства, образование, родной язык, иностранные языки, уровень владения языками (по личной оценке информанта). Все участники указали, что не владеют языком-стимулом.

Реакции респондентов автоматически распознавались как правильные или неправильные, затем происходил ручной контроль неправильных ответов на предмет опечаток. Для автоматического распознавания ответов были составлены списки возможных альтернативных ответов. 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 (см. раздел 4.1).

**3. Орфографическая понятность и социокогнитивные характеристики**

**3.1 Орфографические соответствия**

Установлено, что для 120 болгарско-русских пар когнатов характерны 27 орфографических соответствий (ОС), частотность которых различна, например, *м*:*мл* встречается только один раз в паре когнатов *земя* – *земля*, в то время как ОС *я*:*е* является самым частотным в данном списке когнатов: *вяра* – *вера*, *лято* – *лето*, *сняг* – *снег*, *тяло* – *тело* (см. табл. 1).

**Таблица 1.** Болгарско-русские орфографические соответствия

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№** | **ОС** | **частотность** | **№** | **ОС** | **частотность** |
| 1 | а:я | 3 | 15 | ра:оро | 7 |
| 2 | в:вь | 3 | 16 | ре:ере | 1 |
| 3 | д:дь | 3 | 17 | ре:ерё | 1 |
| 4 | е:ё | 7 | 18 | ръ:ре | 1 |
| 5 | е:о | 6 | 19 | ръ:ро | 1 |
| 6 | е:я | 10 | 20 | т:ть | 9 |
| 7 | и:ы | 13 | 21 | щ:ч | 1 |
| 8 | л:ль | 14 | 22 | ъ:о | 4 |
| 9 | ла:ло | 1 | 23 | ъ:у | 12 |
| 10 | ла:оло | 7 | 24 | ъл:ол | 1 |
| 11 | ле:еле | 1 | 25 | ър:ер | 4 |
| 12 | м:мл | 1 | 26 | ър:ор | 1 |
| 13 | н:нь | 8 | 27 | я:е | 16 |
| 14 | р:рь | 3 |  |  |  |

Очевидно, что различия в орфографии могут повлиять на понимание языка-стимула, но остается вопрос: какие именно вызывают больше проблем у читателя, а какие меньше? Какие орфографические соответствия позволяют носителям русского языка распознавать болгарские слова-стимулы, а какие препятствуют этому (см. раздел 4)?

**3.2 Орфографическая дистанция**

Расчет *орфографической дистанции* между когнатами болгарского и русского языков выполнялся на базе алгоритма, учитываюшего минимальное количество операций вставки или удаления одного символа или же замены одного символа на другой, необходимых для превращения одной строки (цепочки символов) в другую (Levenshtein 1965). В целях нашего исследования алгоритм был модифицирован с учетом лингвистических особенностей, например, согласные буквы можно заменять только согласными, а гласные – только гласными. При этом цена операции вставки, удаления или замены неидентичных символов составляет 1 (например, замена гласной буквы *а* на *о*[[2]](#footnote-2)), а идентичных символов – 0. В табл. 2 представлен в качестве иллюстрации пример расчета расстояния Левенштейна (*Levenshtein distance LD*) для болгарско-русской пары когнатов *младост* – *молодость*, которое в данном примере равняется 3.

**Таблица 2.** Расстояние Левенштейна

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 |

Длина слова играет существенную роль в распознавании слов незнакомого, но (близко)родственного языка (Kürschner, Gooskens, van Bezooijen 2008; Scharpff, van Heuven 1988), причём более длинные слова распознаются лучше, чем короткие. Это, в свою очередь, объясняется соотношением между длиной слова и числом конкурирующих слов-соседей в родном языке, которые очень похожи на слово-стимул, отличаясь от него на один символ (Gooskens 2013; Vitevitch, Rodriguez 2005). Например, у болгарского слова-стимула *младост* (с русским когнатом *молодость,* содержащиморфографическое соответствие *ла*:*оло*) нет орфографических соседей в русском языке, в то время как у болгарского слова *глад* (с русским когнатом *голод* при том же орфографическом соответствии *ла*:*оло*) орфографические слова-соседи существуют: *Влад*,*клад*, *гад*, *глас*, *гладь*, *глаз*, *град*, *лад*[[3]](#footnote-3). Исходная гипотеза заключается в том, что чем длиннее слово-стимул, тем меньше у него орфографических соседей в родном языке и тем легче распознать когнат незнакомого, но (близко)родственного языка.

Полученное расстояние Левенштейна между словами-когнатами *младост*–*молодость* (9) нормализуется по длине цепочки символов следующим образом: 3/9=0,33 или 33%. Нормализованное расстояние Левенштейна (nLD) 33% рассматривается нами как *орфографическая дистанция* и выступает в качестве статистической оценки сложности понимания болгарского слова *младост* носителем русского языка. Итак, с помощью лингвистически модифицированного алгоритма Левенштейна для 120 пар когнатов были рассчитаны орфографические дистанции, которые в дальнейшем соотносились с результатами эксперимента. Данный метрический анализ часто используется в контексте рецептивного мультилингвизма для выявления предсказуемости фонетического и орфографического сходства между когнатами (Gooskens, van Heuven 2019; Möller, Zeevaert 2015): чем больше орфографическая дистанция между соответствующими когнатами, тем сложнее понять слово незнакомого языка и тем больше времени требуется для выполнения задания по свободному переводу (см. раздел 4).

**3.3 Социокогнитивные характеристики**

В процессе проведения эксперимента были собраны метаданные по информантам (пол, возраст, количество языков и пр.). Кроме того, во время выполнения задания фиксировалось время реакции информантов (общее время реакции, начальное время обдумывания стимула, время набора ответа, окончательное время обдумывания ответа). Предполагается, что время, необходимое участнику эксперимента для принятия решения, отражает время обработки и, следовательно, степень сложности задания (Gass, Mackey 2007). Исходя из этого, общее время, необходимое информанту для выполнения задания, должно иметь положительную корреляцию со степенью сложности распознавания стимула. Значительное количество исследований, касающихся времени реакции у мужчин и женщин, свидетельствуют о том, что время реакции у мужчин короче, чем у женщин (Adam et al. 1999; Engel et al. 1972; Der, Deary 2006). Тем не менее Barral и Debû (2004) обнаружили, что, хотя мужчины выполняли задание быстрее женщин, женщины были более точными.

**4. Результаты**

**4.1. Структура выборки информантов**

Как было указано выше, некоторые информанты принимали участие в двух сериях эксперимента, поэтому для нормирования данных было использовано понятие «сессия» – прохождение одной серии эксперимента одним информантом. Структура выборки информантов по полу с количеством пройденных сессий представлена в табл. 3.

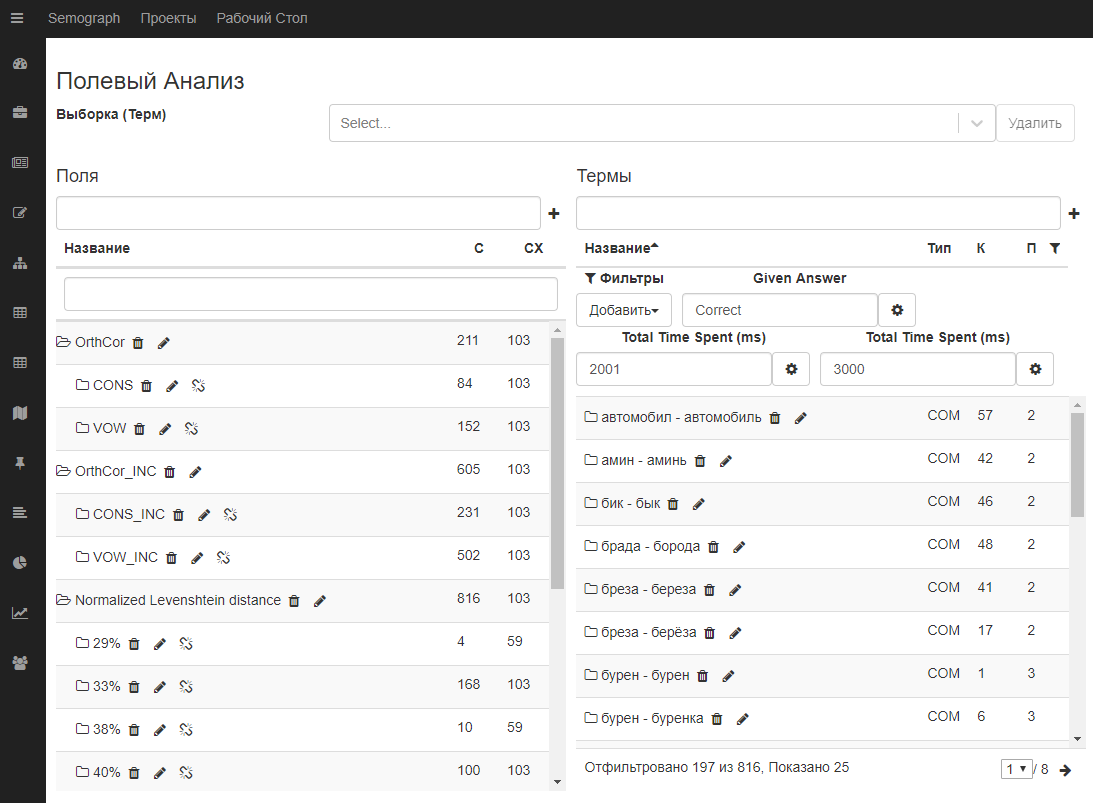
**Таблица 3.** Структура выборки информантов, абс.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Пол | Количество информантов | Количество сессий | Корректных ответов (на сессию), % | Некорректных ответов (на сессию), % | Отсутствие ответов (на сессию), % |
| Женщины | 76 | 89 | 75,30 | 18,86 | 5,84 |
| Мужчины | 27 | 34 | 78,28 | 16,81 | 4,90 |
| Всего | 103 | 123 | 76,79 | 17,84 | 5,37 |

**4.2. Анализ данных**

Полученные в результате эксперимента данные были импортированы в информационную систему «Семограф». Материал был сгруппирован по сессиям. Каждая сессия описывалась, во-первых, метаданными по информанту (пол, возраст, количество языков и др.) и, во-вторых 60 полученными в данной сессии ответами с учетом метаданных, характеризующих каждый из этих ответов (время реакции, корректность ответа и др.). В системе осуществлялась классификация ответов по их корректности, типам орфографических соответствий (ОС) в парах когнатов и орфографической дистанции (nLD, нормализованное расстояние Левенштейна).

На рис. 1 представлена часть окна полевого анализа (классификатора). В столбце «Поля» показан фрагмент многоуровневого классификатора, включающего 3 основных ветви: ОС с корректными ответами (OrthCor), ОС с некорректными ответами (OrthCor INc), nLD. Каждая из ветвей имеет несколько уровней. Так, в ветвь OrthCor входят подуровни CONS (согласные) и VOW (гласные), которые в свою очередь включают конкретные ОС, а они – конкретные пары когнатов.

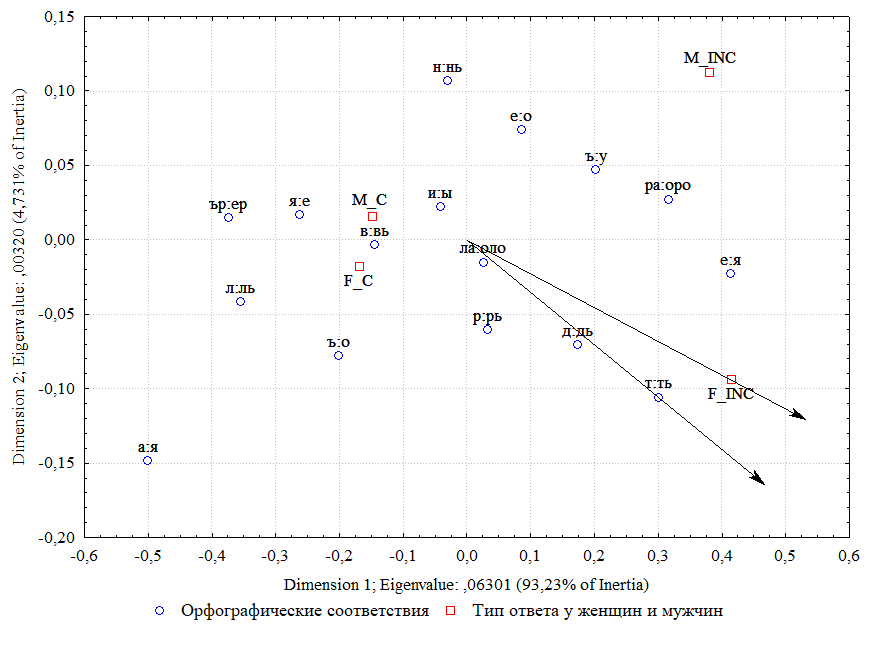


**Рис. 1.** Фрагмент окна полевого анализа в ИС «Семограф»

Благодаря многопараметрической системе фильтрации можно создавать выборки материала по любым метаданным (на рис. 1 представлена выборка ответов с временем реакции от 2001 до 3000 ms).

**4.3. Орфографические соответствия и тип ответа**

Для сопоставления распределений данных разных типов был использован метод анализа соответствий (Correspondence Analysis), который позволяет визуально представить исследуемые данные в координатном пространстве переменных малой размерности, в частности на плоскости (см. Borovikov 2003: 561–576). Система автоматически рассчитывает используемые для анализа соответствия таблицы сопряженности по всему массиву данных или по выборкам из него. Анализ соответствий проведен с помощью программы Statistica 8.0.

На рис. 2 представлены результаты применения метода анализа соответствий к таблице сопряженности «орфографическое соответствие : тип ответа у женщин и мужчин». 

**Рис. 2.** Тип ответа у мужчин и женщин и орфографическое соответствие

Для понимания графика нужно учитывать, что расстояние между данными одного типа свидетельствует о силе связи между ними (чем меньше расстояние, тем связь сильнее, чем больше – тем слабее); сила связи между данными разных типов устанавливается на основе размера угла между двумя точками (относящимися к разным типам данных), вершина которого расположена в центре тяжести графика (точке пересечения осей): острый угол свидетельствует о положительной корреляции (чем меньше угол, тем выше корреляция), тупой – об отрицательной корреляции, прямой – об отсутствии корреляции (Borovikov 2003: 570–571).

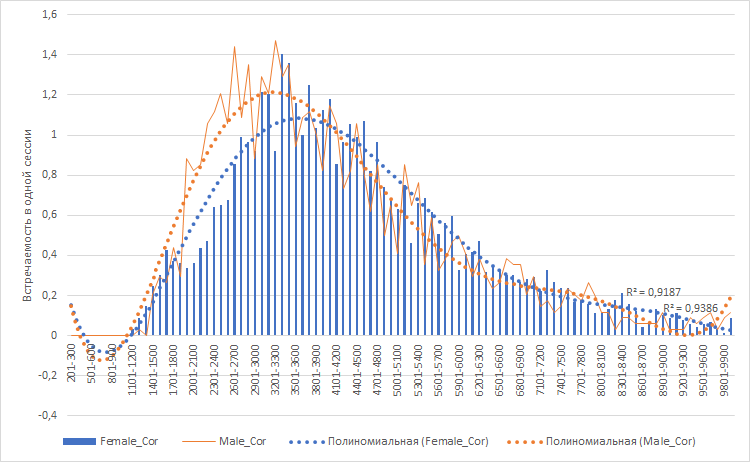
Результаты анализа по всему массиву орфографических соответствий и типов ответа у мужчин и женщин показали, что ось OX (на которую приходится свыше 93% инерции) может быть интерпретирована как шкала, фиксирующая  распределение корректных (слева) и некорректных (справа) ответов. Положение каждого ОС по этой оси можно интерпретировать в контексте его сложности для поиска корректного перевода: чем левее на плоскости находится ОС, тем больше число правильных ответов для содержащих его когнатов, и наоборот, чем правее находится ОС, тем больше число неправильных ответов или отказов. Согласно полученным данным, во всем массиве ОС наиболее легкими для понимания носителями русского языка являются болгарские стимулы, включающие соответствия *ре*:*ерё*, *ре*:*ере*, *е*:*ё*, *ъл*:*ол*, *а*:*я*; самыми сложными оказались стимулы с ОС *ле*:*еле*, *ръ*:*ро*. Все указанные соответствия (кроме *а*:*я*) встретились в экспериментальном материале только в одной паре когнатов каждое (см. табл. 2), т.е. они тесно связаны с конкретными словами, использованными в качестве стимулов. В связи с этим ОС, представленные единичными парами когнатов, были удалены из анализируемого массива (на рис. 2 показаны результаты анализа только для ОС, встречающихся в нескольких парах когнатов).

Ось OY можно проинтерпретировать как шкалу, фиксирующую разницу по полу (хотя доля инерции по данной оси составляет всего 4,73%). Рис. 2 показывает, что корректные ответы у мужчин (M\_C) и женщин (F\_C) находятся рядом, что свидетельствуeт об отсутствии значимых различий, т.е. корректные ответы мужчин и женщин относятся к одним и тем же ОС (и женщины, и мужчины одинаково хорошо переводят когнаты с соответствиями *л*:*ль*, *в*:*вь*, *ър*:*ер*, *я*:*е*, *а*:*я*). Некорректные ответы (M\_INC и F\_INC), напротив, относятся к разным ОС: женщины хуже справляются с когнатами, включающими соответствия *т*:*ть*, *д*:*дь*, *р*:*рь*, *ла*:*оло*, *е*:*я*; мужчины – с когнатами, содержащими соответствия *ъ*:*у*, *е*:*о*. В целом наблюдается тенденция, что для женщин большую трудность при переводе представляют собой когнаты, содержащие орфографические соответствия "согласный : мягкий согласный", а для мужчин – когнаты с парами гласных.

**4.3. Время реакции информантов**

Время поиска языкового соответствия является значимым когнитивным фактором. В данном случае анализировалось общее время реакции информанта от начала демонстрации стимула до сигнала информанта о готовности к переходу к другому стимулу. Анализ времени реакции потребовал устранения статистических выбросов. Для решения этих задач были проанализированы индивидуальные распределения времени реакции для каждого информанта. Для устранения выбросов использовался критерий, сформулированный M.J. Yap и C.S. Seow (2013): из выборки исключаются значения менее 200 ms, а также значения, выходящие за пределы интервала М±2,5SD (индивидуальное среднее ± 2,5 индивидуальных стандартных отклонения).

Вариационный размах временных характеристик (исключая выбросы) был разделен на интервалы по 100 ms. Анализ корректных ответов у мужчин и женщин, нормализованный относительно количества сессий, показан на рис 3.



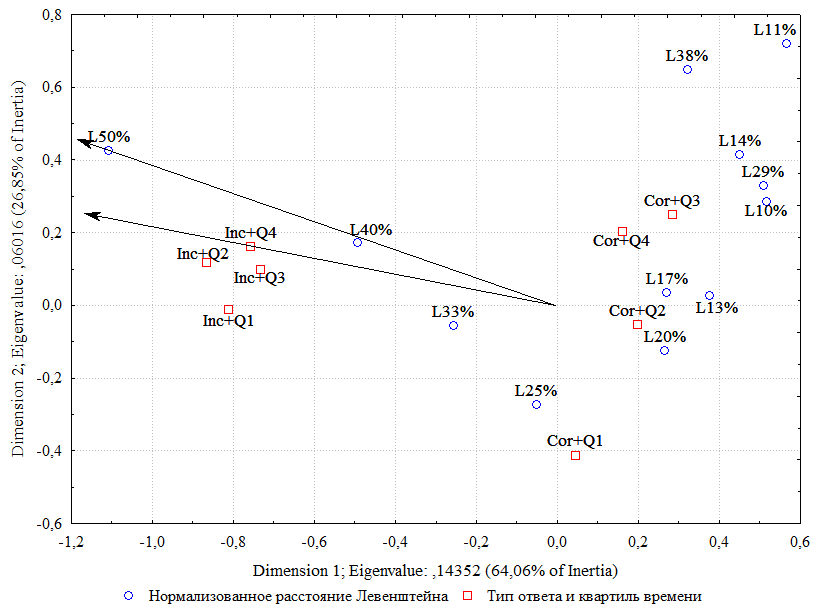
**Рис. 3.** Распределение корректных ответов у мужчин и женщин

по интервалам с шагом 100 ms

На рис. 3 видно, что для корректных ответов требуется не менее 1000 ms; эксцесс полиномиальной функции распределения у мужчин приходится на интервал от 3001 до 3300 ms, а у женщин – на интервал от 3501 до 3800 ms, при этом распределение правильных ответов женщин более пологое, а распределение частот правильных ответов мужчин сдвинуто влево относительно распределения частот правильных ответов женщин, что свидетельствует в целом о более высокой скорости решения данного вида языковой задачи мужчинами.

**4.4. Время реакции и орфографическая дистанция**

Для устранения различий в темпах прохождения эксперимента разными информантами индивидуальные распределения временных характеристик каждой сессии были разделены на квартили; дальнейший анализ проводится на основе агрегирования индивидуальных распределений по квартилям. На рис. 4 визуализированы результаты применения анализа соответствий к таблице сопряженности «квартиль времени реакции корректных и некорректных ответов : нормализованное расстояние Левенштейна».



**Рис. 4.** Нормализованное расстояние Левенштейна и

квартиль времени реакции с типом ответа

Видно, что ось OX (64% инерции) может быть интерпретирована как шкала, фиксирующая  распределение корректных (справа) и некорректных (слева) ответов. Шкала связана и с нормализованным расстоянием Левенштейна: некорректные ответы «захватывают» пары когнатов с диапазоном от 33 до 50%, корректные ответы – диапазон от 11 до 30% (исключение составляет показатель 38%, который характеризует только одну пару когнатов *здраве* – *здоровье*). Это в принципе подтверждает исходную гипотезу, согласно которой степень понятности изолированных когнатов при чтении находится в прямой зависимости от величины орфографической дистанции между когнатами, рассчитанной на базе алгоритма Левенштейна. Ось OY (около 27% инерции) не связана с какими-либо закономерностями распределения данных, хотя, вероятно, она должна быть связана с параметрами нормализованного расстояния Левенштейна (nLD). Этот факт свидетельствует об ограничениях в использовании данного вида метрики к исследованиям с анализом времени реакции.

**5. Заключение**

Экспериментальное исследование с носителями русского языка, осуществленное на веб-платформе INCOMSLAV, показало интересные результаты, связанные с пониманием незнакомого, но родственного языка при чтении. Хотя болгарский язык относится к южнославянской группе, а русский – к восточнославянской, результаты проведенных экспериментов свидетельствуют о том, что для носителей русского языка характерен достаточно высокий уровень спонтанного понимания болгарских стимулов вне контекста – свыше 76% в двух сериях эксперимента.

С помощью информационной системы «Семограф» были проанализированы взаимосвязи между таким лингвистическим фактором как орфографическая понятность (основанным на орфографических соответствиях и орфографической дистанции) и социокогнитивными характеристиками информантов (пол и время реакции). Результаты анализа свидетельствуют о том, что при выполнении задания по свободному переводу болгарских стимулов время реакции у мужчин короче, чем у женщин. При этом и точность ответов у мужчин в среднем выше, чем у женщин. Эти данные не совпадают с оценкой времени ответов мужчинами и женщинами по скандинавским языкам (Schüppert, Gooskens 2011). В этой связи возникает новая постановка вопроса, требующая дальнейших исследований, а именно: следует ли из полученных нами результатов, что женщины склонны дольше обдумывать решение с учетом имеющихся слов-соседей?

Нормализованное расстояние Левенштейна согласно многочисленным исследованиям в большей или меньшей степени адекватно отображает понимание незнакомого, но (близко)родственного языка на орфографическом и фонетическом уровнях. Анализ наших данных также свидетельствует о том, что в целом чем больше орфографическая дистанция между когнатами, тем сложнее понять слово языка-стимула. Однако прямой связи между нормализованным расстоянием Левенштейна и временем реакции обнаружено не было. Возможно следует анализировать время от начала демонстрации стимула до начала ввода ответа.

В целом эксперименты в области рецептивного мультилингвизма при расширении охвата языков и количества информантов позволят решить задачу поиска адекватной системы метрик для оценки близости когнатов. В дальнейшем планируется проводить исследования и с другими парами языков, учитывая влияние дополнительных лингвистических факторов, например, орфографических соседей, частотности слов-когнатов родного языка, а также влияние контекста на распознование слов незнакомого, но родственного языка. Планируется включить в анализ и дополнительные социокогнитивные параметры информантов, например, возраст и знание других иностранных языков. Исследования такого рода важны не только для понимания процессов рецептивного мультилингвизма, но и для развития социолингвистики и лингвистики в целом, а также для методики и дидактики преподавания иностранных языков.

**Литература**

Adam J., Paas F., Buekers M., Wuyts I., Spijkers I., Wallmeyer P. (1999), Gender differences in choice reaction time: evidence for differential strategies, Ergonomics, 42, pp. 327– 335.

Barral J., Debû B. (2004), Aiming in adults: Sex and laterality effects, Laterality: Asymmetries of Body, Brain and Cognition, 9, 3, pp. 299–312.

Borovikov V. (2003), STATISTICA. The art of data analysis on a computer: for professionals, [Iskusstvo analiza dannych na komp’yutere: dlya profissionalov], Piter, SPb. [Piter, SPb.].

Braunmüller K., Zeevaert L. (2001), Semicommunication, receptive multilingualism and related phenomena. A bibliographical overview, [Semikommunikation, rezeptive Mehrsprachigkeit und verwandte Phänomene. Eine bibliographische Bestandaufnahme], Working papers in multilingualism [Arbeiten zur Mehrsprachigkeit], Series B, No. 19, University Hamburg [Universität Hamburg].

Der G., Deary I.J. (2006), Age and sex differences in reaction time in adulthood: Results from the United Kingdom health and lifestyle survey, Psychology and Aging, 21(1), pp. 62–73.

Doyé P. (2005), Intercomprehension, Guide for the Development of Language Education Policies in Europe: From Linguistic Diversity to Plurilingual Education (Reference Studies), Council of Europe, Strasbourg, pp. 105–123.

Engel B.T., Thorne P.R., Quilter R.E. (1972), On the relationship among sex, age, response mode, cardiac cycle phase, breathing cycle phase, and simple reaction time, Journal of Gerontology, 27, pp. 456–460.

Fischer A., Jágrová K., Stenger I., Avgustinova T., Klakow D., Marti R. (2015), An orthography transformation experiment with Czech–Polish and Bulgarian–Russian parallel word sets. In Sharp B., Lubaszewski W., Delmonte R. (eds.), Natural Language Processing and Cognitive Science 2015 Proceedings, Libreria Editrice Cafoscarina, Venezia, pp. 115–126.

Gass S., Mackey A. (2007), Data Elicitation for Second and Foreign Language Research, Erlbaum, Mahwah NJ.

Golubović J., Gooskens C. (2015), Mutual intelligibility between West and South Slavic languages, Russian Linguistics 39, Springer, DOI 10.1007/s11185-015-9150-9, pp. 351–373.

Gooskens C. (2006), Linguistic and extra-linguistic predictors of Inter-Scandinavian intelligibility, In van de Weijer J., Los B. (eds.), Linguistics in the Netherlands 23, John Benjamins, Amsterdam, pp. 101–113.

Gooskens C. (2013), Experimental methods for measuring intelligibility of closely related language varieties, In Bayley R., Cameron R.,  Lucas C. (eds.), Handbook of Sociolinguistics, Oxford University Press, Oxford, pp. 195–213.

Gooskens C., van Heuven V.J. (2019), How well can intelligibility of closely related languages in Europe be predicted by linguistic and non-linguistic variables? Linguistic Approaches to Bilingualism, available at: https://doi.org/10.1075/lab.17084.goo.

Gooskens C., Swarte, F. (2017), Linguistic and extra-linguistic predictors of mutual intelligibility between Germanic languages, Nordic Journal of Linguistics 40(2), pp. 123–147.

HaugenE. (1966), Semicommunication: The language gap in Scandinavia, Sociological Inquiry 36, pp. 280–297.

Kürschner S., Gooskens C., van Bezooijen R. (2008), Linguistic determinants of the intelligibility of Swedish words among Danes, International Journal of Humanities and Arts Computing 2(1/2), pp. 83–100.

Levenshtein V.I. (1965), Binary codes capable of correcting deletions, insertions, and reversals [Dvoichnye kody s ispravleniem udaleniy, vstavok i zamen simvolov], Doklady of the Soviet Academy [Doklady Akademii Nauk SSSR], 1965, Vol. 163, No. 4, pp. 845–848.

Lyashevskaya O.N., Sharov S.A. (2009), Frequency dictionary of the contemporary Russian language [Chastotnyj slovar’ sovremennogo russkogo yazyka], Moskva: Azbukovnik.

Möller R., Zeevaert, L. (2015), Investigating word recognition in intercomprehension: Methods and findings, Linguistics 2015 53(2), pp. 313–352.

Reichert A. (2013), Auditive intercomprehension between Russian and Bulgarian – degree and factors for the listening comprehension of native speakers. In Besters-Dilger J., Schöller U., Slavic language contact, Contributions to the workshop of the same name for students and PhD students, Freiburg, April 19–20, 2013 [Slavischer Sprachkontakt. Beiträge zum gleichnamigen Workshop für Studierende und Promovierende, Freiburg, 19.–20. April 2013], available at: https://freidok.uni-freiburg.de/dnb/download/9232, pp. 119–130.

Scharpff P.J., van Heuven V.J. (1988), Effects of pause insertion on the intelligibility of low quality speech. In Ainsworth W.A., Holmes J.N. (eds.), Proceedings of the 7th FASE/Speech 88 Symposium (Edinburgh), pp. 261–268.

Schüppert A., Gooskens C. (2011), Investigating the role of language attitudes for perception abilities using reaction time. Dialectologia. Special issue, II, pp. 119–140.

Stenger I., Avgustinova T., Marti R. (2017), Levenshtein distance and word adaptation surprisal as methods of measuring mutual intelligibility in  reading comprehension of Slavic languages. In Selegej V.P. (ed.) Computational Linguistics and Intellectual Technologies: International Conference ‘Dialogue 2017’ Proceedings, Issue 16(23), vol. 1, pp. 304–317.

Stenger I., Jágrová K., Fischer A., Avgustinova T., Klakow D., Marti R. (2017), Modeling the Impact of Orthographic Coding on Czech-Polish and Bulgarian-Russian Reading Intercomprehension, Nordic Journal of Linguistic 40(2), pp. 175–199.

Vitevitch M.S., Rodriguez E. (2005), Neighborhood density effects in spoken word recognition in Spanish, Journal of Multilingual Communication Disorders 3(1), pp. 64–73.

Yap M. J., Seow C. S. (2013), The influence of emotion on lexical processing: Insights from RT distributional analysis, Psychonomic Bulletin & Review, 21(2), рр. 526–533.

**Приложение 1 (первая серия эксперимента)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **болгарский** | **русский** | **ОС** | **болгарский** | **русский** | **ОС** |
| автомобил | автомобиль | л:ль | мебел | мебель | л:ль |
| брада | борода | ра:оро | мед | медь | д:дь |
| бурен | бурьян | р:рь, е:я | месо | мясо | е:я |
| видра | выдра | и:ы | мраз | мороз | ра:оро |
| вълк | волк | ъл:ол | мъж | муж | ъ:у |
| вяра | вера | я:е | никел | никель | л:ль |
| вятър | ветер | я:е, ър:ер | овес | овёс | е:ё |
| глад | голод | ла:оло | осел | осёл | е:ё |
| грах | горох | ра:оро | патрул | патруль | л:ль |
| гъсеница | гусеница | ъ:у | пет | пять | е:я, т:ть |
| дим | дым | и:ы | плаж | пляж | а:я |
| дъб | дуб | ъ:у | плесен | плесень | н:нь |
| дъжд | дождь | ъ:о, д:дь | прът | прут | ъ:у |
| един | один | е:о | път | путь | ъ:у, т:ть |
| езеро | озеро | е:о | радост | радость | т:ть |
| жал | жаль | л:ль | риба | рыба | и:ы |
| звяр | зверь | я:е, р:рь | ряпа | репа | я:е |
| земя | земля | м:мл | семе | семя | е:я |
| зет | зять | е:я, т:ть | сняг | снег | я:е |
| злато | золото | ла:оло | сол | соль | л:ль |
| календар | календарь | р:рь | сребро | серебро | ре:ере |
| клас | колос | ла:оло | старост | старость | т:ть |
| кобила | кобыла | и:ы | съпруга | супруга | ъ:у |
| коктейл | коктейль | л:ль | теме | темя | е:я |
| крава | корова | ра:оро | тил | тыл | и:ы |
| кръв | кровь | ръ:ро, в:вь | тяло | тело | я:е |
| кръст | крест | ръ:ре | филм | фильм | л:ль |
| лакът | локоть | ла:ло, ъ:о, т:ть | хлад | холод | ла:оло |
| лен | лён | е:ё | хрян | хрен | я:е |
| мащеха | мачеха | щ:ч | цел | цель | л:ль |

**Приложение 2 (вторая серия эксперимента)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **болгарский** | **русский** | **ОС** | **болгарский** | **русский** | **ОС** |
| амин | аминь | н:нь | месец | месяц | е:я |
| бик | бык | и:ы | младост | молодость | ла:оло, т:ть |
| бреза | берёза | ре:ерё | морков | морковь | в:вь |
| вишна | вишня | а:я | музика | музыка | и:ы |
| врана | ворона | ра:оро | мъка | мука | ъ:у |
| върба | верба | ър:ер | небе | небо | е:о |
| глава | голова | ла:оло | ноздра | ноздря | а:я |
| гняв | гнев | я:е | огън | огонь | ъ:о, н:нь |
| грях | грех | я:е | орел | орёл | е:ё |
| гърло | горло | ър:ор | пелин | полынь | е:о, и:ы, н:нь |
| ден | день | н:нь | пес | пёс | е:ё |
| диня | дыня | и:ы | прах | порох | ра:оро |
| дъбрава | дубрава | ъ:у | пустиня | пустыня | и:ы |
| еж | ёж | е:ё | пъп | пуп | ъ:у |
| език | язык | е:я, и:ы | пяна | пена | я:е |
| елен | олень | е:о, н:нь | пясък | песок | я:е, ъ:о |
| жлеза | железа | ле:еле | ред | ряд | е:я |
| здраве | здоровье | ра:оро, в:вь | резултат | результат | л:ль |
| зъб | зуб | ъ:у | ръка | рука | ъ:у |
| зърно | зерно | ър:ер | син | сын | и:ы |
| име | имя | е:я | славей | соловей | ла:оло |
| келнер | кельнер | л:ль | смърт | смерть | ър:ер, т:ть |
| козел | козёл | е:ё | сряда | среда | я:е |
| коляно | колено | я:е | съпруг | супруг | ъ:у |
| корен | корень | н:нь | сяра | сера | я:е |
| кост | кость | т:ть | хмел | хмель | л:ль |
| крило | крыло | и:ы | цвят | цвет | я:е |
| култура | культура | л:ль | щавел | щавель | л:ль |
| лебед | лебедь | д:дь | яйце | яйцо | е:о |
| лято | лето | я:е | ясен | ясень | н:нь |

1. На данный момент экспериментальная веб-платформа: http://intercomprehension.coli.uni-saarland.de переведена на 11 славянских языков. Участие в экспериментах приняли около 1600 носителей славянских языков (31.01.2019). [↑](#footnote-ref-1)
2. Замена русской буквы ё на болгарскую букву е алгоритмом оценивалась как 0,5. [↑](#footnote-ref-2)
3. Анализ орфографических соседей для каждого болгарского слова-стимула выполнялся с помощью алгоритма Левенштейна на базе лемм (начальных форм) Нового частотного словаря русской лексики (Lyashevskaya, Sharov 2009). [↑](#footnote-ref-3)