

Влияние места словесного ударения на распознавание слов в русской устной речи

Processing initial-stress and non-initial-stress words in spoken-word recognition in Russian

Фёдорова О. В. (olga.fedorova@msu.ru),
Шаврыгина А. С. (shavrygina@gmail.com)

Московский государственный университет им. М. В. Ломоносова

Рассматривается вопрос о принципиальной возможности использования в русском языке (как языке с сильной редукцией) нового экспериментального приема, предложенного в работе [Mattys, Samuel 2000] для проверки метрической стратегии сегментации устной речи. Результаты двух экспериментов дают положительный ответ на поставленный вопрос.

1. Введение. Проблема сегментации устной речи

Одним из важных вопросов в области восприятия и распознавания звучащей речи является вопрос о ее сегментации. Почему при восприятии речи на родном языке эта процедура деления речевого потока на отдельные слова в большинстве случаев происходит быстро и как бы сама собой, а при восприятии незнакомого нам языка превращается в большую проблему? Почему родная речь порой воспринимается нами как слишком медленная, с длинными и четкими паузами между словами, а иностранная часто сливается в один непрерывный поток, ведь в обоих случаях обязательные для письменной речи пробелы между словами при этом отсутствуют? Какие механизмы стоят за этой почти автоматической способностью сегментировать родную речь?

Подобные вопросы, связанные с сегментацией устной речи, исследуются как в области восприятия устной речи, так и в области ее распознавания².

¹ Работа выполнена при частичной финансовой поддержке Российского гуманитарного научного фонда в рамках проекта № 08-04-00165а.

² В отличие от английского термина **speech perception**, который в большинстве случаев переводится на русский язык как восприятие устной (или звучащей) речи, перевод на русский язык термина **spoken word recognition** имеет множество вариантов: устное опознание слова, опознавание слова со слуха, узнавание слова, идентификация слова, распознавание слова на слух, устное распознавание слова, распознавание звучащей речи, распознавание устной речи, распознавание слов устной речи и под. В настоящей работе мы в дальнейшем будем использовать русскоязычные термины **восприятие уст-**

Существующее разделение этих двух научных областей хотя и имеет под собой серьезные основания (традиционно, при изучении восприятия устной речи исследуется, как человек воспринимает и идентифицирует отдельные звуки языка, а при изучении распознавания устной речи нас в первую очередь интересует вопрос идентификации целых слов), но в некоторой степени все же искусственно: нельзя утверждать, что сначала мы распознаем все звуки, а потом складываем из них слова. Наоборот, знание конкретного слова помогает нам правильно распознать звуки, из которого это слово состоит; кроме того, часто нам удается распознать слово еще до того, как оно было произнесено до конца.

Рассмотрим теперь, как вопрос о сегментации решается разными авторами при разных подходах к восприятию и распознаванию устной речи, которые могут быть условно разделены на подходы с опорой на акустические характеристики звуков или на их статистические свойства (так называемые **до-лексические** подходы) и на **лексические** подходы, авторы которых предполагают, что при установлении границ между словами уже активно используется информация о самих этих словах.

ной речи и распознавание (слов) устной речи. В свою очередь, термины **word recognition** и **lexical access** не представляют особой трудности для перевода (**распознавание слов** и **доступ к слову**, соответственно), однако, их точное определение и взаимозаменяемость представляют определенные проблемы. Так, по мнению некоторых авторов доступ к слову происходит раньше распознавания слова, по мнению других — наоборот, позже; во многих работах эти термины используются как синонимы. Мы в дальнейшей работе будем использовать только термин распознавание слова.

В свою очередь, лексические подходы представлены, в частности, когортной моделью Марслен-Вильсона [Marslen-Wilson 1973] (ее последняя версия [Gaskell, Marslen-Wilson 2002], однако, содержит некоторые элементы коннекционизма), а также несколькими коннекционистскими моделями, среди которых стоит выделить TRACE [McClelland, Elman 1986] и Shortlist [Norris 1994].

Остановимся немного более подробно на до-лексических теориях, связанных с дистрибутивными (или статистическими) характеристиками слов. Одна из основополагающих идей при этом подходе состоит в том, что при определении границ между словами слушающий активно опирается на просодические свойства слов, а именно, что каждый сильный слог является потенциальным началом нового слова. Согласно данной **стратегии метрической сегментации** речи [Cutler & Norris 1988] слова с ударением на первом слоге обрабатываются быстрее и легче, чем слова с ударением не на первом слоге. Этот факт был подтвержден уже достаточно большим количеством экспериментальных исследований, однако в работе [Mattys, Samuel 2000] авторам впервые удалось подготовить стимулы таким образом, чтобы использовать один и тот же акустический материал в различных экспериментальных условиях.

2. Экспериментальная методика из работы [Mattys, Samuel 2000]

Эксперимент, описанный в работе [Mattys, Samuel 2000], проводился по методике **обнаружения фонемы** (phoneme monitoring или phoneme detection [Connine, Titone 1996]): испытуемому называют определенный целевой звук (или показывают на экране соответствующую букву), который ему нужно будет распознать, после чего он слушает предложения или список несвязанных между собой слов и должен нажать на определенную клавишу, как только услышит целевой звук. Считается, что испытуемые выполняют подобные задания не на до-лексическом уровне обработки, а уже на лексическом. В качестве контроля выполнения задания именно на лексическом уровне эксперимент часто дополняется вспомогательным чисто лексическим заданием (в частности, в эксперименте, описанном в работе [Mattys, Samuel 2000], испытуемый должен был нажимать на специальную клавишу всякий раз, когда слышал слово, относящееся к категории орудий труда).

В каждой экспериментальной попытке испытуемый слышал последовательность из семи слов, произнесенных практически без пауз, и должен был нажать на клавишу, как только услышит целевой звук. Этот целевой звук всегда был начальным согласным слога и мог находиться в (i) начальном

безударном слоге, (ii) начальном ударном слоге, (iii) втором безударном слоге и (iv) втором ударном слоге. Особенностью данного эксперимента было то, что слова с разным местом ударения в действительности представляли собой один и тот же акустический материал, то есть одни слова были вырезаны из других. Так, исследователи записывали пару слов-близнецов (test twins) *saga-zealous*, а потом получали из нее четыре комбинации: /g/ во втором безударном слоге слова *saga* (вариант iii), /g/ в начальном безударном слоге слова *gazelle* (вариант i), /z/ в начальном ударном слоге слова *zealous* (вариант ii) и /z/ во втором ударном слоге слова *gazelle* (вариант iv). Таким образом авторам работы [Mattys, Samuel 2000] удалось избежать проблем, связанных с вариативностью произносимых слов.

3. Эксперименты на русском материале

Вышеописанная стратегия метрической сегментации (а также более строгий её вариант, который гласит, что только слова с главным ударением на первом слоге воспринимаются как начала слов и благодаря этому требуют меньше времени и когнитивных усилий для своей обработки) проверялась по большей части на материале английского языка, где 90% всех слов имеют ударение на первом слоге (кроме английского тестировался также голландский язык [Vroomen, de Gelder 1995]). Таким образом, проверка данной гипотезы на русском материале не только добавит в исследовательскую копилку еще один язык, но и даст возможность верифицировать эту теорию на принципиально ином метрическом материале.

Однако проведение аналогичных экспериментов на русском материале³ может осложниться сильной редуccionией, характерной для русского языка по сравнению с английским. Слова, вырезанные подобным образом из русских слов-близнецов, могут вызвать у испытуемых серьезные проблемы с их распознаванием. Поэтому прежде чем приступать к проведению эксперимента, аналогичного описанному в [Mattys, Samuel 2000], нам предстояло провести большую подготовительную работу по проверке потенциальной возможности распознавания слов, вырезанных из слов-близнецов. Описанию двух таких экспериментов и будет посвящена следующая часть нашей работы.

³ Фонетические, фонологические и просодические особенности русского языка, несомненно, очень хорошо изучены (свежий обзор психолингвистических работ см., например, в [Ягунова 2008]), однако работ по аналогичной [Mattys, Samuel 2000] методике «вырезания» слов, насколько нам известно, на материале русского языка еще не проводилось.

3.1. Эксперимент с использованием методики регистрации движений глаз

В качестве стимульного материала нами было составлено 24 пары слов-близнецов (например, *доминотарелка*, полный список см. в Приложении 1), 12 из вырезанных слов были хорейми, 12 остальных — ямбами. Затем слова-близнецы были записаны и обработаны в звуковом редакторе Cool Edit Pro таким образом, чтобы между словами практически не было пауз, после чего были составлены и записаны (тоже без пауз на стыках слов) «семерки» слов (24 собственно экспериментальных, 12 отвлекающих (филлеров) и 4 тренировочные). Все слова, входящие в экспериментальные семерки, были среднечастотными, имели привычное звучание на стыках; семерки были записаны по такому принципу, что позиция целевого слова в них была сбалансирована от 2-го до 6-го места; две пары слов в семерке были записаны вместе (слова в скобках), а две пары слов отдельно, например: *картон — нота — (бамбук — карандаш) — школьник — (ковёр — радость)*; на каждом из двух экспериментальных листов семерки с хорейским целевым словом чередовались с семерками с ямбическим целевым словом. В экспериментальных листах филлеры и экспериментальные семерки стояли на тех же самых позициях, но если на одном листе целевое слово было «вырезанным», то на другом листе на том же самом месте стояло целиком записанное целевое слово.

В процессе тестирования экспериментального материала мы провели несколько пилотных экспериментов, в которых задавали испытуемым вопрос: «Есть ли в данной последовательности слов что-то не очень естественно звучащее?». Большинство испытуемых указывали в этом случае на записанные вместе пары слов и на наши экспериментальные вырезанные слова. Таким образом, гипотеза, которую мы проверяем в ходе этого эксперимента звучит так: «Будет ли тот факт, что слова, вырезанные из двух слов-близнецов, звучат не совсем естественно, мешать их быстрому и правильному распознаванию по сравнению со словами, записанными целиком?».

Эксперимент был выполнен в виде презентаций в программе Microsoft PowerPoint. Инструкция испытуемым была следующей: «В каждой экспериментальной попытке Вы будете слышать последовательность из семи слов, начитанных с уменьшенными паузами между словами (что затрудняет их понимание); одновременно на экране Вы будете видеть восемь слов, только одно из которых будет совпадать с услышанными Вами словами. Ваша задача как можно быстрее кликнуть мышью по совпадающему слову. Через секунду после этого вокруг правильного слова (независимо от того, правильным или нет был Ваш ответ) будет появляться красная рамка».

Данный эксперимент был проведен с шестью испытуемыми по методике записи свободных движений глаз (free-viewing eye-tracking, оборудование ETL-500

фирмы ISCAN Inc., подробнее об этой методике см. [Федорова 2008]), при этом фиксировались момент воспроизведения ключевого слова, момент первого взгляда на ключевое слово, момент нажатия на мышку, правильность выбора, а также высчитывались время с момента воспроизведения ключевого слова до первого взгляда на ключевое слово (наиболее важный параметр) и время с момента воспроизведения ключевого слова до нажатия мышки. Результаты (см. Приложение 2) свидетельствуют о том, что статистически значимое различие есть только во времени ответов на вырезанные и записанные целиком ямбические слова.

3.2. Эксперимент на воспроизведение слов

В нашем втором эксперименте были использованы те же 40 экспериментальных семерок, но каждая из них воспроизводилась два раза подряд, чтобы исключить пропуск целевого слова в том случае, если с первого раза испытуемые не смогут запомнить все слова. В инструкции говорилось, что эксперимент имеет целью определение объёма рабочей памяти в условиях плохой перцептивной различимости слов. После прослушивания стимульной семерки испытуемые должны были воспроизвести все слова, которые они смогли запомнить; сразу после первой попытки слова данной семерки воспроизводились второй раз, что помогало некоторым испытуемым повторить во второй попытке большее количество слов. Если в первом эксперименте мы смотрели, будет ли в случае вырезанных слов увеличиваться время, необходимое на поиск целевого слова, и количество неправильных ответов, то в этом эксперименте мы проверяли гипотезу «будет ли тот факт, что слова, вырезанные из двух слов-близнецов, звучат не совсем естественно, мешать их правильному воспроизведению по сравнению со словами, записанными целиком?».

Эксперимент, представленный в виде презентаций в программе Microsoft PowerPoint, был проведен с 40 испытуемыми. Результаты (см. Приложение 3) говорят о том, что (i) повторение в семерках-филлерах (в среднем 5 из 7) статистически значимо лучше, чем в экспериментальных словах (4,15 из 7); (ii) повторение в **целиком записанных словах** (4,3 из 7) статистически незначимо лучше, чем в вырезанных (4 из 7); (iii) повторение целевого слова в **целиком записанных хорейх** статистически незначимо лучше, чем в вырезанных хорейх; (iv) повторение целевого слова в **целиком записанных ямбах** статистически значимо лучше, чем в вырезанных ямбах; (v) повторение целевого слова в **хорейх** статистически значимо лучше, чем в ямбах.

В последнем разделе настоящей работы мы сопоставим результаты двух экспериментов, подведем итоги и кратко наметим перспективы дальнейших исследований.

4. Обсуждение результатов

Итак, по результатам двух проведенных экспериментов можно сделать общий вывод, что в целом не совсем естественное звучание вырезанных слов не мешает ни их правильному распознаванию (эксперимент 1), ни их правильному воспроизведению (эксперимент 2). Таким образом, русская редукция не препятствует принципиальной возможности проведению эксперимента, аналогичного [Mattys, Samuel 2000], на русском материале.

Обратим, однако, внимание на тот факт, что и в первом и во втором экспериментах вырезанные ямбические слова и распознавались хуже (время реакции в первом эксперименте статистически значимо больше), и воспроизводились хуже (количество правильно воспроизведенных слов для вырезанных ямбов значимо меньше как по сравнению с вырезанными хорейми, так и по сравнению с записанными целиком ямбами, в то время как для хореев это различие статистически незначимо). Более того, в целом все (т.е. и вырезанные, и записанные целиком) хорейские слова во втором эксперименте воспроизводились значимо лучше, чем все ямбические слова. Данный факт, по нашему мнению, может свидетель-

ствовать как о несовершенстве нашего ямбического стимульного материала, так и о том, что хорейские слова русского языка понимаются (то есть распознаются и воспроизводятся) лучше, чем ямбические. Последняя гипотеза может сама по себе уже служить некоторым подтверждением использования не только англоязычными, но и русскоязычными носителями стратегии метрической сегментации.

Для того, чтобы развеять эти две гипотезы, нам нужно будет провести еще один подготовительный эксперимент, аналогичный эксперименту 2, заменив в стимульном материале все слова, особенно ямбические, которые плохо распознавались и воспроизводились по результатам уже проведенных экспериментов (см. Приложение 4). Если при замене наиболее неудачных слов (как то: ямбические слова *топор, коза, комар, роса, нога* и *коса*; хорейские *палец* и *ветка*), мы получим аналогичные результаты, это будет уже с большей вероятностью свидетельствовать в пользу использования в русском языке стратегии метрической сегментации. Однако более строго доказать это утверждение можно будет только после проведения основного эксперимента, подобного описанному в работе [Mattys, Samuel 2000], на русском материале.

Приложение 1. Список экспериментальных слов-близнецов

№	хорей	целевое слово	звуки	№	ямб	целевое слово	звуки
1	домино-тарелка	нога	н-т	1	слово-дама	вода	в-д
2	письмо-река	море	м-р	2	мясо-ваза	сова	с-в
3	такси-лопата	сила	с-л	3	полка-маршал	комар	к-м
4	молоко-жара	кожа	к-ж	4	конфета-порция	топор	т-п
5	тропа-лицо	палец	п-л	5	палка-занавес	коза	к-з
6	ответ-карман	ветка	в-к	6	ворона-галстук	нога	н-г
7	слеза-поход	запах	з-п	7	долина-радуга	нора	н-р
8	попугай-конверт	гайка	г-к	8	телега-радио	гора	г-р
9	метро-залив	роза	р-з	9	утро-сабля	роса	р-с
10	толпа-русалка	парус	п-р	10	берёза-водка	завод	з-в
11	борода-чапушка	дача	д-ч	11	колесо-сокол	носок	н-с
12	кошмар-лягушка	марля	м-л	12	точка-сахар	коса	к-с

Приложение 2. Первый эксперимент

Первая цифра — целиком записанные слова, вторая через запятую — вырезанные.

	правильные ответы (из 36)	время
хорей	33, 29	1.302, 1.447
ямб	32, 26	1.327, 2.191

Приложение 3. Второй эксперимент

Первая цифра — целиком записанные слова, вторая через запятую — вырезанные.

	кол-во слов (из 1680)		кол-во целевых (из 240)	
	1 попытка	2 попытка	1 попытка	2 попытка
хорей	1067, 1028	1294, 1257	132, 114	180, 159
ямб	1007, 936	1266, 1139	101, 42	155, 72
филлер	2419 (из 3360)	2814 (из 3360)		

Приложение 4. Стимульные слова

Первая цифра — целиком записанные слова, вторая через запятую — вырезанные).

Слова, выделенные жирным шрифтом, распознавались плохо; слова, выделенные курсивом, распознавались на среднем уровне.

№	слово	тип ударения	первый эксперимент, ошибки	первый эксперимент, время	второй эксперимент, целевое слово
1	нота	хорей	0, 1	долго, долго	9, 12
2	сова	ямб	0, 1	быстро, быстро	17, 12
3	море	хорей	0, 0	быстро, быстро	13, 12
4	топор	ямб	1, 1	быстро, быстро	6, 3
5	сила	хорей	0, 1	быстро, быстро	4, 11
6	коза	ямб	1, 0	быстро, быстро	6, 2
7	кожа	хорей	0, 1	быстро, быстро	9, 5
8	вода	ямб	0, 0	быстро, долго	8, 1
9	палец	хорей	1, 0	быстро, долго	1, 5
10	комар	ямб	0, 2	быстро, быстро	5, 1
11	ветка	хорей	0, 2	быстро, быстро	17, 0
12	<i>нора</i>	<i>ямб</i>	0, 1	быстро, быстро	10, 2
13	запах	хорей	0, 0	быстро, быстро	9, 11
14	роса	ямб	1, 1	быстро, быстро	5, 0
15	гайка	хорей	0, 1	быстро, быстро	15, 11
16	завод	<i>ямб</i>	0, 0	быстро, быстро	10, 2
17	роза	хорей	1, 0	долго, быстро	13, 14
18	нога	ямб	0, 1	быстро, быстро	7, 1
19	парус	хорей	0, 0	быстро, быстро	13, 20
20	<i>гора</i>	<i>ямб</i>	0, 1	долго, быстро	8, 3
21	дача	хорей	0, 0	быстро, быстро	17, 8
22	коса	ямб	1, 1	быстро, быстро	8, 1
23	<i>марля</i>	<i>хорей</i>	0, 1	быстро, долго	12, 5
24	носок	ямб	0, 1	быстро, быстро	11, 14

Литература

1. Фёдорова О. В. Методика регистрации движений глаз «Визуальный мир»: шанс для сближения психолингвистических традиций // Вопросы Языкознания, 6. 2008. С. 98–120.
2. Язунова Е. В. Вариативность стратегий восприятия звучащего текста. Пермь, 2008.
3. Connine C. M., & Titone D. Phoneme monitoring // Language and Cognitive Processes, 11(6), 1996. P. 635–646.
4. Cutler A. & Norris D. The role of strong syllables in segmentation for lexical access // Journal of Experimental Psychology: Human Perception and Performance, 14, 1988. P. 113–121.
5. Gaskell M. G. & W. D. Marslen-Wilson. Representation and competition in the perception of spoken words // Cognitive Psychology, 45, 2002.
6. Marslen-Wilson W. D. Linguistic structure and speech shadowing at very short latencies // Nature, 244, 1973.
7. Mattys S. & Samuel A. Implications of stress pattern differences in spoken word recognition // Journal of Memory & Language, 42, 2000. P. 571–596.
8. McClelland J. L. & Elman J. L. The TRACE model of speech perception // Cognitive Psychology, 18, 1986.
9. Norris D. Shortlist: a connectionist model of continuous speech recognition // Cognition, 52, 1994. P. 189–234.
10. Vroomen J. & de Gelder B. Metrical segmentation and lexical inhibition in spoken-word recognition // Journal of Experimental Psychology: Human Perception and Performance, 21, 1995. P. 98–108.