

К ВОПРОСУ О ГРАНИЦАХ ПРЕДЛОЖЕНИЯ В УСТНЫХ ТЕКСТАХ НА ЯЗЫКЕ БЕЗ УСТОЙЧИВОЙ ПИСЬМЕННОЙ ТРАДИЦИИ¹

ON SENTENCE BOUNDARIES IN ORAL TEXTS IN LANGUAGES WITHOUT STABLE WRITTEN TRADITION

Захаров Л. М. (leon@philol.msu.ru)

Филологический факультет МГУ им. М. В. Ломоносова

Казакевич О. А. (kazak@orc.ru)

НИВЦ МГУ им. М. В. Ломоносова

В докладе предполагается рассмотреть проблему разграничения предложений в устной речи на языке без устойчивой письменной традиции. Дело в том, что членение дискурса на предложения, столь очевидное в письменной речи, зачастую вызывает затруднения, коль скоро мы имеем дело с устным текстом на языке без устойчивой письменной традиции. На материале практически бесписьменных языков Сибири (селькупского и кетского) будет рассмотрена возможность использования просодических признаков в качестве опоры для членения устного текста на предложения.

1. Членение на предложения первичной графической записи устного текста

Сегодня обычным методом записи устной речи является ее непосредственная фиксация с помощью какого-нибудь звукозаписывающего устройства (например, кассетного или цифрового диктофона), а уже затем перевод аудиозаписи в графическую запись — ее расшифровка, осуществляемая в процессе неоднократного прослушивания сделанной аудиозаписи, как правило, с помощью информантов-носителей языка.

Вопрос о членении текста на предложения возникает уже в процессе графической записи текста в полевых условиях. Однако в полевых условиях этот вопрос отступает на задний план: прежде всего важно записать последовательность произносимых рассказчиком слов, зафиксировать их значение, вычленив идиомы, понять, о чем идет речь в отдельных фрагментах текста, а уж как эти фрагменты соотносятся с нашим представлением о предложении, полученным на базе письменных текстов, можно решить потом. Не удивительно поэтому, что в полевых блокнотах лингвистов границы предложений явно обозначаются далеко не по всему тексту. Обычно решения по вопросу границ приходится принимать на стадии последующей обработки полевых записей: например, подготовки их к публикации или вводу в базу данных.

Еще сравнительно недавно — всего несколько десятков лет назад — далеко не все лингвисты, отправляясь в экспедицию, имели при себе звукозаписывающие устройства. Тексты по большей части фиксировались непосредственно на бумаге и только на ней. Поэтому при последующей обработке текстов решение о границах предложений в неясных случаях могло приниматься в основном на базе семантических критериев, что зачастую предполагало неоднозначность результата. Приведем один, на наш взгляд, небезыntenесный пример.

В конце 1980-х годов формировался компьютерный корпус селькупских текстов. Одной из составляющих корпуса были тексты из архива замечательной исследовательницы селькупского языка Л. А. Варковицкой, записанные в 1941 г. и сохранившиеся в ее полевых блокнотах². Для ввода текстов в компьютер были привлечены студенты ОТиПЛ МГУ — они занимались этим во время летней практики. Студентам была объявлена норма, выполнив которую они получали зачет по практике. Норма устанавливалась в предложениях. В полевых блокнотах Л. А. Варковицкой границы предложения обычно отмечаются стандартно: заглавной буквой в начале и точкой в конце. Однако в некоторых случаях точек не стоит, а заглавные буквы можно отнести на счет имен собственных. В задачу студентов входило, таким образом, не только собственно ввод текста, но иногда и принятие решения насчет границ предложений. Смысл вводимых текстов был студентам понятен, так как в блокнотах под каждой строкой селькупского текста обычно дан русский перевод. В результате мы получили довольно большой разброс в длине предложений корпуса, и этот разброс отражает не только разные стратегии рассказчиков, от которых записаны тексты (что, вообще-то очевидно), но и разные стратегии студентов,

¹ Работа выполнена в рамках проекта «Взаимодействие сегментного и супraseгментного уровней в фонетике языков Сибири (на материале контактирующих языков среднего течения Енисея и сопредельных территорий)», грант РФФИ № 05-06-80234.

² Об архиве Л. А. Варковицкой см. [Кузнецова, Хелимский 1989].

вводивших эти тексты в компьютер. Одни студенты по возможности минимизировали длину вводимых предложений, другие собирали короткие фрагменты в сложные предложения, увеличивая объем собственной работы, но следуя своему представлению о том, как должен быть устроен текст.

Пример (1) представляет сочетание «минималистской» стратегии рассказчика (предложения 1.1–1.3) с «минималистской» стратегией ввода (предложения 1.3–1.4 и 1.5–1.6):

- (1) *Šitty tīm̄nasyq̄aqy* ‘Два брата’. (1.1.)
Qällyt nenty tūññ□#yt ‘Ненцы вместе пришли’. (1.2.)
Okkyr qällyl’ m□#yr ‘Один ненецкий богатырь’. (1.3.)
Sitty näläqyt ‘Две девки его’. (1.4.)
Qällyl’ m□#yr nił’cik k↔tyty ‘Ненецкий богатырь так сказал’. (1.5.)
«Mat sitty näläqy tonty ütentaḡ» ‘«Моих двух дочек к озеру пошлю я»’. (1.6.)

Пример (2) — составление сложносочиненного предложения из простых (2.2).

- (2) *Nyny losy nił’cik k↔tyñyt* ‘Потом черт так сказал.’ (2.1.)
«Ivan tsarevic, me ti to maty na q↔nnynty, mat, tat q□##ly qampisä qottä sorät, ükyl’ ohyl’ tiry nuñentymyt» ‘«Иван-царевич, мы сейчас в тот чум пойдем, я, ты лоб-твой платком завяжи, в шапках мы стоять будем»’.

2. Анализ аудиоматериалов: постановка задачи

Одним из результатов технического прогресса стало неперенное использование лингвистами при сборе материалов устной речи звукозаписывающих устройств. Таким образом, помимо графической расшифровки текстов в руках исследователя оказывается исходное звучание самих этих текстов, и это звучание можно не только прослушивать снова и снова, но и анализировать с помощью компьютерных программ анализа звучащей речи, ставших доступными массовому пользователю.

В грамматиках языков Сибири довольно часто можно найти ссылку на интонацию как показатель границ предложения, однако при этом, как правило, отмечается, что этот вопрос пока не исследован (см., например [Прокофьев 1935: 99; Терещенко 1973: 15; Кузнецова и др. 1980: 398]). Иногда, однако, к использованию интонации для определения границ предложения, напротив, относятся скептически (см. Werner 1997: 343).

На материале селькупских и кетских текстов мы попытались определить, существует ли корреляция между отрезками текста, которые для нас естественно рассматривать как отдельные предложения, исходя из их значения и грамматических характеристик, с одной стороны, и интонацией этих отрезков, с другой. Иными словами, можно ли рассматривать интонацию в качестве критерия при определении границ предложений в кетских и селькупских текстах.

Интонация (фразовая просодия) определяется как «совокупность просодических характеристик предложения: тона, громкости, длительности, фонации» [Кодзасов, Кривнова 2001: 380]. Тон считается важнейшей из этих характеристик. Иногда к интонации относят и паузирование [Маслов 1997: 81]. В проведенном анализе мы учитывали расстановку пауз и движение тона. Анализ проводился с помощью программы Speech Analyzer. Рассматривалась 1) функция паузы в тексте; 2) просодические признаки границы синтагмы и предложения.

3. Анализ кетского материала

Материалом анализа послужили фольклорные и бытовые тексты на суломайском говоре южнокетского диалекта, записанные в 2004 г. в пос. Суломай Байкитского района Эвенкийского АО³. Приводимые ниже примеры взяты из «Сказки о кукушка», записанной от В. Н. Тыгановой (1942 г. р.).

Пауза в текстах выступает как маркер 1) границы синтагмы, 2) границы предложения, 3) хезитации. Наиболее длительные паузы наблюдаются при хезитации. А вот конец синтагмы (в том числе и конечной синтагмы предложения) совсем не обязательно оформляется паузой (см. фразу 5).

Падение тона перед паузой или резкий перепад тона при отсутствии паузы (фраза 5) можно считать маркером границы синтагмы, в том числе и конечной синтагмы предложения. Повышения тона в конце синтагмы в рассмотренном материале не наблюдается. В целом в конце предложения падение тона происходит на большую «глубину», чем в конце неконечной синтагмы, однако существуют и исключения (см. фразу 5). Впрочем, фразу 5 возможно разбить на два отдельных предложения.

³ Экспедиция 2004 г. в Суломай проводилась в рамках проекта «Мультимедийная база данных кетского языка», грант РФФИ 04-04-12028в.

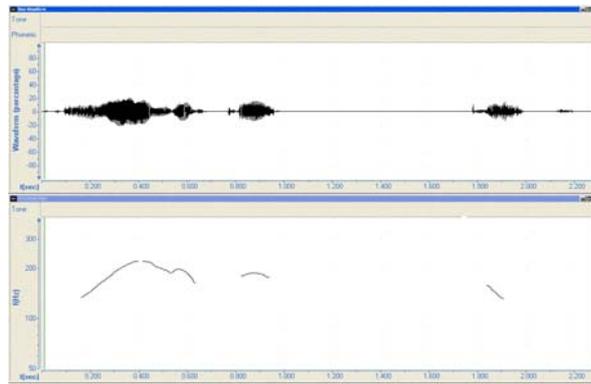


Рис. 1. Фраза 1 (кетский язык)

*D↔o'l'd↔χ qo²q | kε²t ||*⁴

‘Жила одна женщина [человек]?’.

Две синтагмы. Межсинтагменная пауза ~ 160 ms. Перед ней нет резкого падения тона: по интонации слышно, что синтагма неконечная (в отличие от мелодического оформления конечной синтагмы).

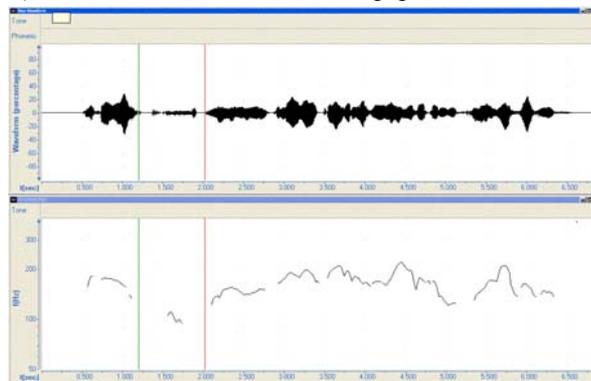


Рис. 2. Фраза 2 (кетский язык)

Bu qa↔j↔|...| naŋa m ϕ↔n↔: «Dil'g↔r↔, k↔ydin ab↔η ul'b↔ś↔↔ten↔itn, ɕa urós baj ↔binśam!» ||

‘Она потом им сказала: «Дети! Идите мне воды хоть принесите, я захотела пить?»’.

Пауза хезитации (заполненная) — отмечена маркерами — ~ 820 ms.

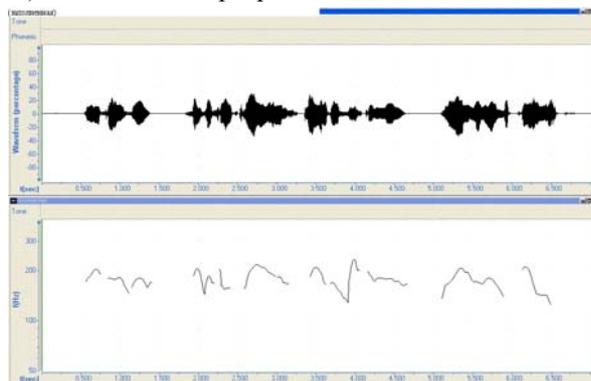


Рис. 3. Фраза 3 (кетский язык)

Tur↔ qaj↔↔bu | dyl'yt qaj↔↔qaujaŋ | ɛla qaj↔↔dusk↔mnim | nā am naŋal' d↔jo ϕ↔joq. ||

‘Это потом она, дети потом зашли, дверь опять открыли, их мать от них улетела’.

Четыре синтагмы. Длительность межсинтагменных пауз: 1 ~ 440 ms, 2 ~ 70 ms, 3 ~ 440 ms. В первой и третьей («физиологических») слышно как диктор набирает воздух для продолжения речепроизводства. Перед межсинтагменными паузами нет резкого падения тона: по интонации слышно, что синтагмы неконечные (в отличие от последней — конечной).

⁴ При записи текстов приняты следующие обозначения:

| — короткая пауза;

|| — длинная пауза;

↔ — движение основного тона вниз;

↔ — движение основного тона вверх;

' — знак, обозначающий динамическое выделение слога (можно условно назвать словесным ударением) — ставится перед «ударным» слогом (для селькупских текстов).

‘Потом говорит: «Один сын, одна дочь!»’

Длительность межсинтагменной паузы перед прямой речью ~ 600 ms. Во втором случае паузы практически нет (отсутствие сигнала вызвано смычкой при произнесении *q*), однако граница отчетливо ощущается интонационным оформлением. Можно отметить также, что просодически схожее оформление конца синтагмы перед прямой речью и в конце фразы перцептивно воспринимается по-разному: если в первом случае слушатель ожидает продолжения, то во втором — продолжение не обязательно. Скорее всего, это можно объяснить передачей рассказчиком интонации персонажа в прямой речи.

4. Анализ селькупского материала

Материалом анализа послужили фольклорные и бытовые тексты на верхнетолькинском и среднетазовском говорах северного диалекта селькупского языка, записанные в 2001 г. в пос. Толька Пуровского района⁵ и в 2002 г. и пос. Красноселькуп Красноселькупского района⁶ Ямало-Ненецкого АО. Приводимые ниже примеры взяты из «Сказки о старике и старухе», записанной от В. К. Аксандакова (1957–2003, среднетазовский говор).

Пауза в селькупских текстах, также как и в кетских, выступает как маркер 1) границы синтагмы, 2) границы предложения, 3) хезитации. Характерной особенностью селькупского нарратива является полное отсутствие повышения тона в конце синтагмы, будь то конечная синтагма в предложении или неконечная — проведенное исследование подтверждает выводы, ранее полученные на другом материале [Захаров, Казакевич 2003]. При этом конечную синтагму от неконечной отличает не столько большая длительность паузы (ее может почти вовсе не быть), сколько значение тона в конце синтагмы: тон в конечной синтагме падает ниже, «глубже» (у диктора, примеры речи которого приведены в докладе, до 80 Гц), чем в неконечной.

Для иллюстрации наших наблюдений позволим себе привести довольно пространный пример, представляющий собою не отдельные фразы, а начало селькупской сказки о жадном и глупом старике и находчивой и сердобольной старухе.

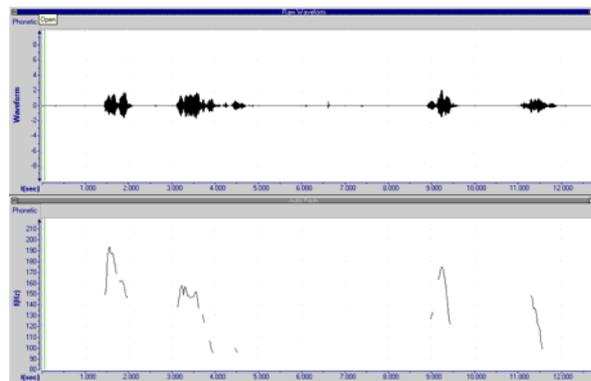


Рис. 7. Фразы 1–2 (селькупский язык)

1. *'ilym ɕ pa | 'ira' imaqo ɕ tantys(a) 'kyn p ɕ rō(qy)t ||*
 ‘Жил старик со своей старухой в верховьях реки’.

2. *na ɕ tangyt || 'k' iŋcyp ɕ 'omtyl' tyjyty. ||*
 ‘В это лето запор поставил’.

Предположительно отличие конечной и неконечной синтагм заключается не в понижении / повышении тона, а в частотном диапазоне, в котором находится значение тона в конце синтагмы. Сравни: в неконечной синтагме в конце синтагмы $F = 146$ Гц (в первом случае) и $F = 121$ Гц (во втором), значение тона для конечной синтагмы $F = 95$ Гц (в первом случае) и $F = 99$ Гц (во втором).

Значение пауз: после неконечной синтагмы в первом случае $t = 990$ мс и $t = 1300$ мс — во втором по сравнению с паузой после конечной синтагмы $t = 3980$ мс (т. е. в 3–4 раза больше). Между 2 и 3 фразами $t \sim 2000$ мс.

⁵ Экспедиция 2001 г. в Пуровский район Ямало-Ненецкого АО проводилась при поддержке РФФИ, грант 01-06-8020.

⁶ Экспедиция 2002 г. в Красноселькупский район Ямало-Ненецкого АО проводилась при поддержке РГНФ, грант 02-04-18019е.

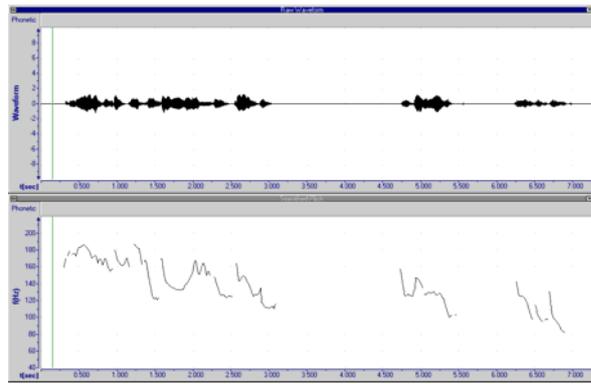


Рис. 8. Фраза 3 (селькупский язык)

3. *'k'ihñöynty cap ɕ' tünta 'caŋal' ɕ' nyrša, 'qōsa || ni 'χal' ni ɕ' jaryx || 'q↔#yty ɕ' cäŋka ||*
 'В запор его лишь приходит только ерш, окунь, никакой другой рыбы-его нет'.

Значение длительности паузы внутри фразы после слова *qōsa* $t = 1680$ мс (т. е. значительно больше, чем среднее значение), по-видимому, единственным критерием того, что синтагма неконечная, является достаточно высокое значение тона в конце синтагмы $F = 110$ Гц по сравнению со значением тона в конце конечной синтагмы $F = 80$ Гц. Длительность второй межсинтагменной паузы $t = 780$ мс. Значение тона в конце синтагмы после слова *jaryx* $F = 100$ Гц (сравни с $F = 80$ Гц в конце фразы).

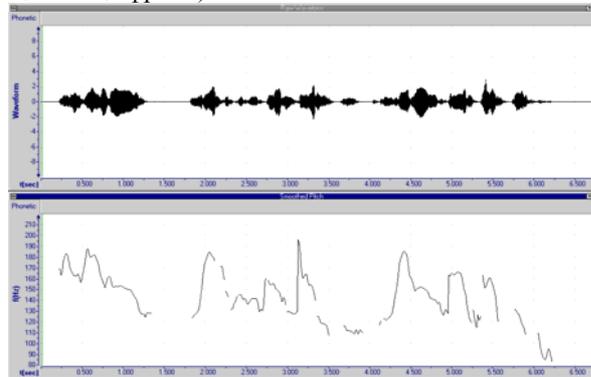


Рис. 9. Фраза 4 (селькупский язык)

4. *'ira 'heññy ɕ' mŋl↔ || na 'mortyty nyny 'konna 'sonnyt na l 'tal l↔#l' oyl' 'q↔#yyp 'qaj 'tattysyty ||*
 'Старик, рассердившись, из котца⁷ на берег черпает это, эту костеголовую рыбу что принесло'.

Длительность первой паузы $t = 530$ мс; длительность второй — $t = 150$ мс.

Значение тона в конце первой синтагмы $F = 125$ Гц; значение тона в конце второй синтагмы $F = 109$ Гц; значение тона в конце третьей (конечной) синтагмы $F = 83$ Гц.

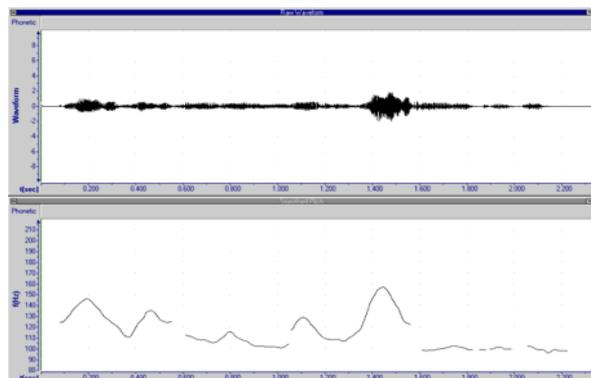


Рис. 10. Фраза 5 (селькупский язык)

5. *'rom 'ilä nyny 'aj 'tonnytyty ||*
 'Палку взяв, потом опять бьет'.

Значение тона в конце $F = 98$ Гц.

7

⁷ «Котец» — часть «запора» — приспособления для ловли рыбы.
 «Костеголовая рыба» — рыба малоценных пород (окунь, ерш и пр.)
 «Чир», «подчир» — ценная рыба, рода сигов, семейства лососевых.

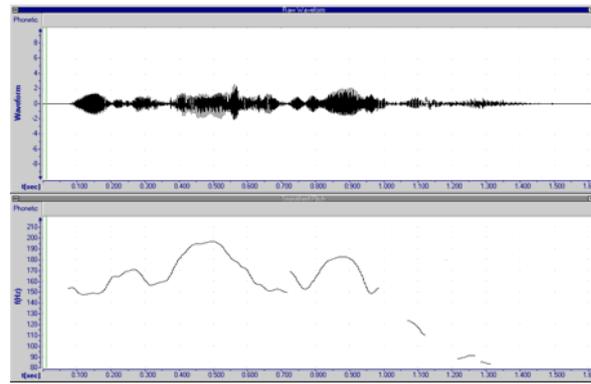


Рис. 11. Фраза 6 (селькупский язык)

6. *nyŋ q↔#yŋ mel'ty ūr'reja* ||
 'Потом рыба совсем потерялась (пропала)'
 Значение тона в конце F = 83 Гц.

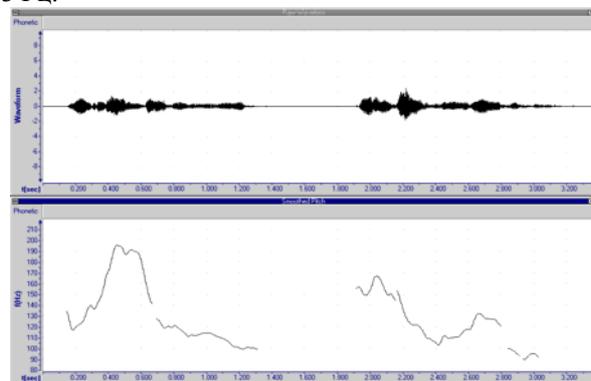


Рис. 12. Фраза 7 (селькупский язык)

7. *ni c'nyršaty, ni | c'χōZaty, ni 'qajty cāŋka* ||
 'ни ерша, ни окуня, ничего нет'.
 Длительность межсинтагменной паузы t = 640 мс.
 Значение тона в конце неконечной синтагмы F = 100 Гц.
 Значение тона в конце конечной синтагмы F = 93 Гц.

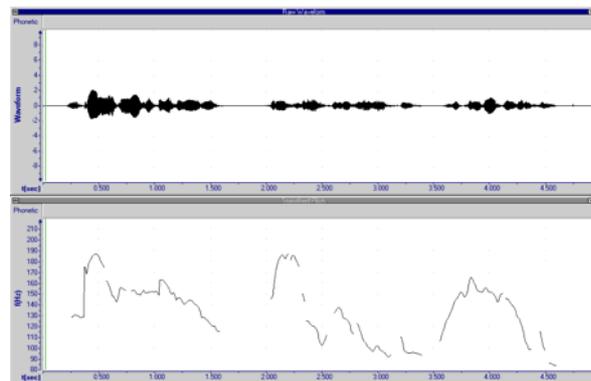


Рис. 13. Фраза 8 (селькупский язык)

8. *na 'tanjnty kunty qapi i | 'icca kiŋ c'ōdqnty* || **никакой** *c'q↔#y cāŋka* ||
 'В течение этого лета (все лето) вроде (постоянно) ходит к запору-своему **никакой** рыбы нет'.
 Длительность первой паузы t = 440 мс; длительность второй паузы t = 190 мс.
 Значение тона в конце первой синтагмы F = 116 Гц; значение тона в конце второй синтагмы F = 93 Гц;
 значение тона в конце третьей синтагмы F = 83 Гц.

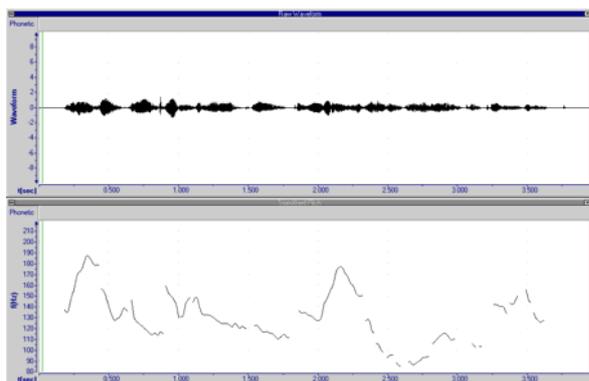


Рис. 14. Фраза 9 (селькупский язык)

9. 'mannypat nyny 'kiš'šēyt ⊆↔χ'q↔#y ⊆'ulDelim⊆Ba ⊆a ⊆'morty ⊆'aš'šēryšty ||
 'смотрит потом возле запора рыба плавает, а в котец не заходит'.

Одна синтагма. Нетипичное движение основного тона: в конце синтагмы $F = 127$ Гц, не в конце синтагмы $F = 85$ Гц.

При этом в конце синтагмы (как обычно) тон падает. Возможно, тон передает интонацию (эмоцию) удивления.



Рис. 15. Фразы 10–11 (селькупский язык)

10. Ну ⊆ладно ||

11. Уже 'aranyl' ⊆'taŋy ⊆na ⊆'taŋyDu Gun'Du kuttar-χos na iŋmyntoj, | 'amnat kuttar aš'χump⊆#qi ||

'Уже осень (осеннее лето — сентябрь месяц, начало октября), этого лета в течение как-то жили (они двое), с голоду чуть не умирали'.

Длительность межсинтагменной паузы в селькупской фразе $t = 196$ мс.

Длительность паузы в между русской (кодировое переключение) и селькупской фразой $t = 523$ мс.

Значение тона в конце русской синтагмы $F = 86$ Гц; значение тона в конце первой селькупской синтагмы $F = 83$ Гц; значение тона в конце последней синтагмы $F = 80$ Гц. То есть по значению тона первая селькупская синтагма — конечная, по длине паузы — неконечная.

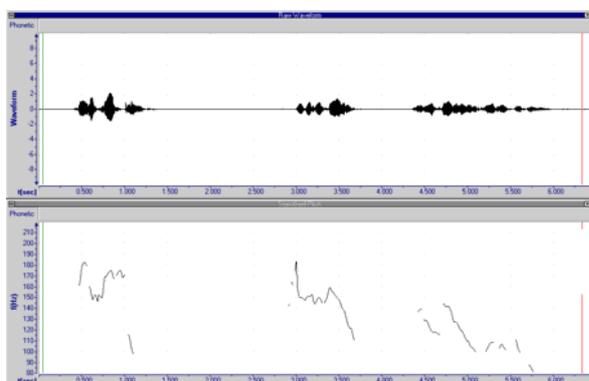


Рис. 16. Фразы 12–13 (селькупский язык)

12. 'syryp 'cat⊆tyŋyt ||

'Снег выпал'.

13. 'ukkur to ⊆'χaryt || 'ira 'kiŋ⊆cōxynty χarra 'kur⊆rolna ||

'Однажды утром старик к забору-своему отправился'.

Длительность паузы между 12 и 13 фразами $t = 1540$ мс.

Длительность межсинтагменной паузы в 13 фразе $t = 590$ мс.

Значение тона в конце 12 фразы $F = 98$ Гц; значение тона в конце первой синтагмы в 13 фразе $F = 110$ Гц; значение тона в конце последней синтагмы $F = 81$ Гц.

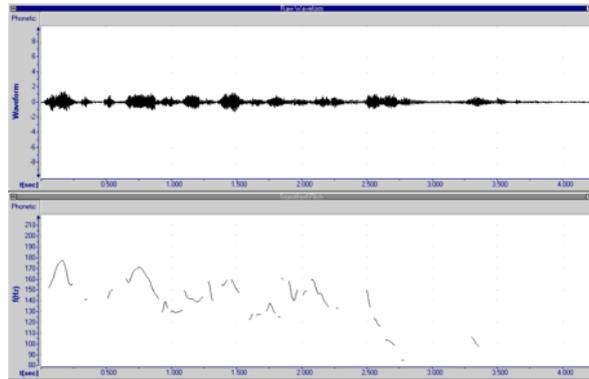


Рис. 17. Фраза 14 (селькупский язык)

14. *'mortykty' qolč̣ṭympat, 'qolč̣ṭympat, ukkur ypč̣' qorija / ɕs̄ōqoltyt ||*
 'В котце черпает, черпает, одного подчира зачерпнул'.

Значение межсинтагменной паузы ~ 300 мс.

Значение тона в конце первой синтагмы 83 Гц. Значение тона в конце конечной синтагмы программа определить не может из-за слишком тихого произнесения.

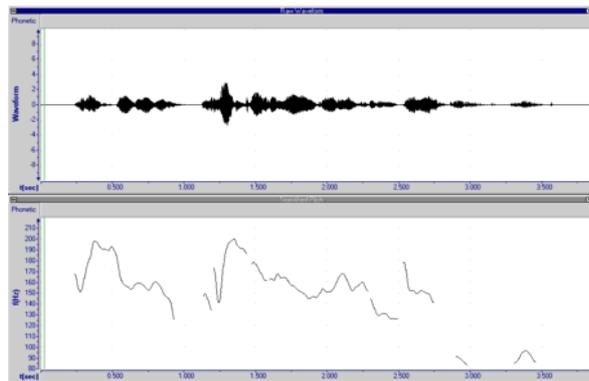


Рис. 18. Фраза 15 (селькупский язык)

15. *'ypkolija, 'ypkolijanty ɕ'w↔ɕɕyla 'konnnä ɕtanta / 'mʷ#χynty ||*
 'Подчира, подчира-своего подняв, на берег вышел к дому-своему'.

Во фразе есть сбой (+ кашель).

Есть маленькая пауза (90 мс) перед последним словом

5. Некоторые выводы

Несмотря на то, что кетский и селькупский языки не состоят даже в отдаленном родстве и кардинальным образом различаются по внутреннему устройству практически на всех языковых уровнях, просодия селькупского и кетского предложения имеет много общих черт. Если можно предположить, что функции паузы в устной речи практически универсальны для любого языка (или, по крайней мере, для большинства языков), то набор возможных интонационных контуров предложения в целом и отдельных синтагм внутри предложения может существенно меняться от языка к языку. Общим для кетского и селькупского языков является практически полное (или почти полное) отсутствие синтагм с восходящим тоном на конце, падение тона в конце **всех** синтагм при минимальных его значениях в конечных синтагмах предложения. При этом в селькупском нарративе эта особенность представлена ярче и четче, чем в кетском, что наводит на мысль о том, что если отмеченное сходство просодии предложения — результат тесных языковых контактов, следует признать, что специфическая интонация заимствовалась из селькупского языка в кетский, а не наоборот. Представляется также, что для селькупского языка (по крайней мере, для рассмотренных говоров северного диалекта) членение устного текста на предложения можно проводить с опорой на интонацию, в то время как для кетского языка подобный алгоритм

менее очевиден⁸. Все это, безусловно, требует дальнейших исследований с привлечением материалов большего числа информантов-носителей разных говоров как с кетской, так и с селькупской стороны.

Список литературы

1. Захаров Л. М., Казакевич О. А. Экспериментальное исследование фразовой интонации при кодовом переключении // *Компьютерная лингвистика и интеллектуальные технологии. Диалог'2003*. М.: Наука, 2003. С. 250–253.
2. Кодзасов С. В., Кривнова О. Ф. *Общая фонетика*. М.: РГГУ, 2001.
3. Кузнецова А. И., Хелимский Е. А. *Селькупские материалы в рукописном наследии Л. А. Варковицкой // Советское финно-угроведение*. 1989. Т. 25. № 1.
4. Кузнецова А. И., Хелимский Е. А., Грушкина Е. В. *Очерки по селькупскому языку. Тазовский диалект. Том 1*. М.: Изд-во МГУ, 1980.
5. Маслов Ю. С. *Введение в языкознание*. М.: Высшая школа, 1997.
6. Прокофьев Г. Н. *Селькупский (остяко-самоедский язык)*. Л., 1935
7. Терещенко Н. М. *Синтаксис самодийских языков*. Л.: Наука, 1973.
8. Werner H. *Die ketische Sprache*. Wiesbaden: Harrassowitz Verlag, 1997.

⁸ Что, по-видимому, и отражено в противоположных взглядах на роль интонации в определении границ предложения, представленных у исследователей селькупского и шире — самодийских языков, с одной стороны, и кетского языка, с другой (см. отсылки в разделе 3).