

**ПАУЗАЦИЯ В ЯПОНСКОМ ЯЗЫКЕ НА ГРАНИЦАХ
СИНТАКСИЧЕСКИХ ЕДИНИЦ РАЗНОГО УРОВНЯ:
КОРПУСНОЕ ИССЛЕДОВАНИЕ¹**

**PAUSES ON THE DIFFERENT TYPES OF SYNTACTIC BOUNDARIES
IN JAPANESE: A CORPUS STUDY**

Комарова А.Д. (komarovichka@gmail.com)

Российский государственный гуманитарный университет

Данное исследование посвящено паузам на границе синтаксических единиц разного уровня в устной японской монологической речи. Его цель состоит в том, чтобы проследить, насколько частотны (и вероятны) паузы на границах предложений и меньших синтаксических единиц (клауз), а также какова их «нормальная» длина.

Данная работа представляет собой корпусное исследование на материале устных монологических рассказов на японском языке.

Ее цель заключается в том, чтобы проследить, с какой частотой в устной японской речи возникают паузы на границах синтаксических единиц разного уровня и какова может быть стандартная длина этих пауз. Кроме того, для каждого типа границ прослеживается соотношение абсолютных пауз и пауз заполненных (как лексически, то есть, элементами, имеющими словарное толкование типа русского «ну»), так и долексически, то есть единицами типа «мм»), так как заполнение пауз часто может указывать на затруднения говорящего при порождении текста.

Для анализа использовался корпус из тридцати трех текстов, «рассказов по картинкам». От каждого информанта записывалось по четыре текста, однако от двоих информантов из-за плохого качества записи использовались неполные комплекты (по три и два текста). Тексты записывались в два приема. Сначала записывались рассказы, а через 6-8 часов – пересказы. Носителям предлагались два набора картинок с несложным сюжетом. Просмотрев одну серию картинок, информанты составляли по ней рассказ, имея картинки перед глазами. Информанты были предупреждены, что эксперимент состоит из двух частей, но им не сообщалось, в чем именно заключается вторая часть. При записи пересказов картинок они не видели и восстанавливали сюжет по памяти.

Паузация представляет собой неотъемлемую часть устного дискурса, однако долгое время при изучении языка за основу бралась письменная форма, и такие распространенные в устном дискурсе явления, как, например, заполненные паузы, ассоциировались только с хезитацией и нарушениями плавности речи.

Во второй половине двадцатого века появились работы, рассматривающие паузацию в связи с синтаксическими границами. Так, в [Chafe 1994] вводится понятие «интонационных единиц» и показывается, что в 60% случаев в английском языке они соответствуют клаузам.

В [Кривнова, Чардин 1999] описывается паузирование на границах «интонационных фраз» в русском языке. Авторы отмечают, что паузы используются для передачи синтаксических и смысловых отношений. Характерно, что большинство границ предложений в русском языке отмечается паузами, и паузы часто совпадают с границами простых предложений в составе сложного.

Данные японского языка по устному дискурсу представлены не столь хорошо, как русского и английского, однако существует ряд работ, посвященных этой теме. Так, [Iwasaki 1993] отмечает, что интонационные единицы в японском языке, как правило, мельче, чем в английском. Однако, [Sakura, Fuji 2006] показали, что большинство интонационных единиц соответствуют клаузам, а исключения составляют топики, более характерные для японского, чем для английского языка.

Данное исследование посвящено паузации на границах синтаксических единиц двух типов в японском языке:

1. Границы предложений
2. Границы между клаузами, не являющиеся границами предложений (т.е. в общем случае границы между простыми предложениями в составе сложного)

¹Работа выполнена при финансовой поддержке РФФИ, грант 07-06-00061а

Сегментирование устного дискурса – одна из важных проблем, возникающих при транскрибировании. Если на письме автор текста как правило использует пунктуационные знаки для обозначения синтаксических границ, то в устном дискурсе подобных средств нет. Вместо графических символов используются просодические характеристики: так, на наше понимание текста влияют интонация и наличие пауз.

Как и письменный, устный дискурс членится на предложения, а также меньшие предикации в составе сложных предложений. Сегментирование текста не всегда тривиально, поэтому в данной работе будут рассматриваться только стопроцентные границы клауз. Такие границы выделялись на основании синтаксического критерия.

В первую очередь, это границы между предложениями. Как правило, конец предложения маркируется адресивными формами глагола на *masu/desu*. Большинство предложений корпуса оканчивается глаголом в адресивной форме. Всего в корпусе 301 предложение. Из них 227, т.е. 75,4%, оканчивается на адресивные формы глагола. Вот пример такого предложения:

(1) DAI T-2

1. **Kyou wa ... (0.8) e (0.7) oyasumi des-u.**
сегодня TOP HON-выходной день COP.ADR-PRS
Сегодня выходной день.

Исключение составляют заголовки (пример (2)) и подытоживающие фразы с семантикой «конец», где может не быть финитного глагола. Всего таких случаев в корпусе 10 (что составляет 3,3 %). Новым также считается предложение, начатое с обрыва (см пример (7)). В корпусе всего 12 обрывов (т.е. 4% от числа предложений).

(2) WAK R-1

1. **Nimaime no ue, ... (0.6) dainikkaime.**
второй лист GEN согласно второй раз
Второй лист, второй раз.

Отдельную проблему представляет собой вычленение предложений, которые оканчиваются на неадресивные формы глагола. Так, в японском языке глагол, употребленный в «простой», т.е. финитной неадресивной форме, может употребляться как сказуемое или как определение. Часто симптомом конца предложения в корпусе может являться частица, следующая после простой формы, как в примере (3):

(3) ISI T-1

6. **... (0.9) Baggu .. (0.3) ranpu ... (0.7) ee (0.4) to nabe ni**
сумка лампа HEZ кастрюля DAT
.. (0.2) kabin iroiro na mono ga ar-u na (0.2).
ваза разный ATR вещь NOM иметься-PRS PRT
Есть разные вещи: сумки, лампы, кастрюли, вазы.

В целом же только один из рассказчиков предпочел для двух из своих текстов простые формы и их функция очевидна из контекста. Вот такой отрывок:

(4) ISI T-1

9. **kime-rare-na-i**
решать-POT-NEG-PRS
не могу решить.
10. **Soo da.**
так COP.PRS
Так вот.
11. **soko ni kodomo-tachi ga i-ru no**
там DAT ребёнок-PL NOM быть-PRS NML
kii-te mi-yoo
спрашивать-CNV AUX.M-HOR
Попробую спросить у детей, которые здесь есть.

В 9-ой строке употреблена простая форма, однако она не может быть понята, как определение к 10-ой строке, поскольку 10-ая строка не начинается с именной группы. То же относится к 11-й строке.

Не сигнализируют о границе предложения адресивные формы глагола, употребленные с противительным союзом *ga*:

(5) WAK T-1

21. **.... (2.2) Totemo yosasoo na kuruma ga ar-imash-ita ga,**
очень хороший ATR машина NOM иметься-ADR-PST но

Паузация в японском языке на границах синтаксических единиц разного уровня

22.(0.6) *nedan o kik-u to,*
 цена ACC спросить-PRS TEMP
 Хотя и была очень хорошая машина, (когда) спросил цену...

Есть также 1 пример неканонического конца предложения:

(6) SHU T-1

8. *o-too-san anmashi o-kane ga na-kute*
 HON-отец-сан почти HON-деньги NOM иметь.NEG-CNV
 9. ..(0.2) *ka-e-na-i to*
 купить-POT-NEG-PRS выходит так
 (так как) у отца почти нет денег, выходит так, что купить он не может.

В данном случае информанты утверждают, что это конец предложения. Возможно, что финитный глагол после союза (видимо, цитационного) *to* опущен.

Кроме границ предложений, «стопроцентными» границами клауз можно считать границы после:

1. конвербов на *te/de* (не в составе аналитических форм глагола)
2. конвербов на *i* (совпадающих со второй основой глагола и так же не в составе аналитических форм глагола)
3. союзов (условно-временного *to*, соединительного *shi*, противительных *ga*, *keredo(mo)*, уступительного *noni*, причинного *node* и т. п.); союзных имен (например, *toki*) и следующих за ними падежных послелогов (таких, как *ni*) и частиц (например, *wa*); послелогов в составе предикатной группы (*ni*, *kara*, *made* и др). А также граница клауз проводится ПЕРЕД цитационным союзом *to* с глаголами мышления и говорения (e.g. *to ui*, *to kangaeu*, *to omou*...) Наряду с цитационным союзом рассматриваются цитационные конструкции *to/sou iu fuu ni*

1. Паузы на границах предложений

Итак, по данным корпуса на границе предложений почти всегда - т.е. почти в 94 % случаев - возникают паузы. Из них в 23% случаев это заполненные паузы.

Вот редкий случай, когда паузы на границе предложений нет:

(7) KEN T-1

38.(9.2) *Sorede(0.3) sooyu= sooyu koto o ==*
 тогда вышеназванный NML ACC
 39. *Uchi e modot-te kara(0.3),*
 дом в возвращаться-CNV после
 после возвращения домой...

В приведенном примере видно, что говорящему не удается породить то предложение, которое устроило бы его, и поэтому он бросает неудачный сегмент и начинает новый. При этом времени на планирование у него довольно много: перед забракованным отрывком он делает длительную паузу. К тому моменту, когда он решает отменить неудавшийся отрывок и начать предложение заново, он, видимо, уже знает, что именно хочет сказать.

Вот еще один пример, когда в начале предложения нет паузы (см. начало 3-й строки):

(8) MAY T-2

2. ..(0.3) *asagohan o tabe-mash-ita.*
 завтрак ACC есть-ADR-PST
 3. *Sono ato ni ...(0.8) ee(0.6)to sukii ni dekake-mash-ita.*
 это после DAT HEZ лыжи DAT отправляться-ADR-PST
 Позавтракал. После этого отправился кататься на лыжах.

В данном случае говорящий, предположительно, заранее спланировал несложный отрезок. Второе предложение совсем короткое и, в принципе, говорящий мог бы слить эти два предложения в одно. Кроме того, второе предложение начинается союзным наречием, которое представляет собой эксплицитный показатель связи и также могло повлиять на длину (в данном случае - отсутствие) паузы.

В целом же можно говорить о том, что обычно в позиции границы предложения паузы возникают. Эти паузы дают говорящему возможность сделать вдох и спланировать следующий отрывок, а слушающему наряду с другими просодическими характеристиками служат сигналом о наличии синтаксической (как правило, обусловленной семантическими связями) границы.

Что касается «нормальной» длины пауз на границе предложений, то это очень нестрогий параметр. Длина пауз может варьироваться в зависимости от темпа речи говорящего, жанра текста, более или менее знакомой говорящему тематики и т.д.

Среднее арифметическое не может считаться вполне надежным показателем, т.к. одна слишком затянутая пауза или, напротив, несколько коротких предложений с незначительными паузами на границе могут сильно повлиять на это значение.

Поэтому для определения «стандартной» длины пауз на границах предложений в устном неподготовленном рассказе использовались, кроме среднего арифметического, такие показатели, как мода и медиана. Безусловно, сами эти показатели варьируются в разных текстах, однако среднее значение медианы для длины пауз на границе предложений в данном корпусе составило 1,4 секунды, и самая частотная длина паузы на границе предложений в корпусе (т.е. мода) совпадает с этим показателем.

2. Паузы на границах клауз

2.1. Паузы после конвербов на *te/de*

Конвербы на *te/de* – наиболее частотная инфинитная форма глагола, которая встречается в японской речи. Всего в корпусе 136 клауз, заканчивающихся на данную форму. И приблизительно в 55% случаев на таких границах возникают паузы. Из них в 15 % случаев эти паузы – заполненные. При этом данный показатель сильно варьируется для разных текстов (тенденция может соблюдаться для говорящего в целом): так, говорящий может вообще не делать пауз после форм на *te/de*, как в тексте KEN T-2 или, напротив, отграничивать паузой каждую клаузу на *te/de*, как в тексте MAY T-1.

Вот пример границы клаузы на *te/de* без паузы:

(9) WAK R-2

- | | | | |
|----|--|--|--|
| 7. | ..(0.2) sukii o hai-te, | | |
| | лыжи ACC надеть-CNV | | |
| 8. | dokoka e dekake-te ik-imash-ita | | |
| | куда-то в выйми-CNV AUX.DIR-ADR-PST | | |
- Надев лыжи, куда-то вышел.

И с паузой:

(10) SHU T-2

- | | | | | |
|----|--|--|--|--|
| 7. |(1.3) yopparat-ta mama sikii o shi-te | | | |
| | пьянеть-PST так.как.есть лыжи ACC заниматься-CNV | | | |
| 8. | ...(0.9) koron-de, | | | |
| | упасть-CNV | | | |
- [будучи] пьяным катаясь на лыжах, упал.

Средняя длина пауз после данной формы составляет около 0,5 секунды, медиана - 0,6 секунды, а мода – 0,2 секунды.

2.2. Конверб на *-i*

Формы на *-i* встречаются в текстах значительно реже, чем формы на *te/de*. Всего в тестах встретилось 24 клаузы, заканчивающихся на данные формы. В 16 из 24 случаев после них были паузы, четверть из них (всего четыре паузы) были заполненными. Таким образом, процент пауз после форм на *-i* несколько выше, чем после форм на *te/de*, и составляет 67%, однако выборка слишком мала и не дает оснований утверждать, что паузы после форм на *-i* вообще частотнее.

Что касается таких параметров, как мода, медиана и среднее значение, то они также весьма условны. Медиана и среднее значение очень близки и составляют приблизительно 0,6 секунды. Мода равна 0,4 секунды.

Примеры клауз, оканчивающихся конвербом на *-i*:

(11) KEN T-2

- | | | | |
|----|----------------------------------|--|--|
| 7. | sukii ita o moch-i, | | |
| | лыжи DO брать | | |
| 8. | sukii jou ni dekake-masu. | | |
| | трасса ALL выходит-ADDR.PRS | | |
- Позавтракав. (он) взял лыжи и вышел на трассу.

(12) MAI R-2

- | | | | |
|----|---|--|--|
| 3. | ...(0.5) ee(0.2) choushoku o tab-e, | | |
| | завтрак ACC есть-CNV | | |
| 4. | ...(0.5) de sorekara ..(0.1) sukii ni dekake-mash-ita. | | |
| | затем после.этого лыжи DAT отправляться-ADR-PST | | |
- Позавтракав, отправился кататься на лыжах..

2.3. Паузы на стопроцентных границах клауз, кроме клауз, заканчивающихся конвербами на *te/de* и *-i*

Кроме границ клауз, которые оканчиваются на глагольные формы *te/de* и *-i*, в корпусе встретилось еще 159 стопроцентных границ клауз. Это границы после союзов и послелогов в составе предикатной группы, а также границы перед цитационными союзами.

В корпусе встретились следующие показатели:

Паузация в японском языке на границах синтаксических единиц разного уровня

показатель	вхождений в корпусе
ato «после»	9
to-TEMP «если/когда»	14
kedo «хотя»	6
node «так как»	24
-tara «если»	5
toka «или»	5
kara «после»	5
ga «но»	25
keredomo «несмотря.на»	3
tameni «благодаря»	1
niyotte «в.зависимости»	1
-tari репрезентатив	1
ni (в том числе <i>toki ni, sugu ni, mae ni</i> , 2-я основа+ni, 1-я основа+ zu+ni) DAT с разными оттенками значения	10
toki (+toki wa) «во.время»	2
-shi «и»	2
-nagara «делая»	1
sei de «из-за»	2
mata «как.есть»	1
-zu отрицание	1
-ba «если»	1
-temo «даже если»	1
dakara «так как»	1
to QUOT цитационный союз	36

Таблица 1

Процент пауз на таких границах достаточно высок и составляет в среднем около 63% . Из них в 18% случаев эти паузы являются заполненными.

(13) WAK T-1, 22

22. *nedan o kik-u to*
цена ACC спросить-PRS TEMP

23. *totemo taka-i node,*
очень высокий-PRS так.как

24. *..(0.4) ka-u koto ga deki-mas-en*
купить-PRS NML NOM мочь-ADR-NEG

(когда) спросил цену, поскольку (она была) очень высокая, купить не смог.

(14) ISI T-2

1. *...(0.6) tokoroga ..(0.4) sukii ga ..(0.3) hm ...(0.8) eto*
но лыжи SUBJ

sakamichi ni sashikakat-ta toki ni
уклон IO зацепиться-PAST время TEMP

2. *..(0.3) toma-ru koto wa deki-zu ni*
останавливаться-INF SBST TOP мочь-NEG FOC

3. *..(0.3) Watasi wa ..(0.3) aaa(0.5) ...(0.6) koron-de shimai-mash-ita.*
Я TOP падать-COP заканчиваться-ADDR-PAST

Но, когда лыжи приблизились к склону, я не смог их остановить и упал.

В (13) примере иллюстрируется употребление условно-временного союза *to* и причинного *node*. При этом после *to* паузы нет, а после *node* небольшая 0.4 сек.пауза есть. В примере (14) приводятся случаи употребления союзного имени и отрицательной формы глагола. В обоих случаях на границе клауз возникают паузы.

Среднее значение длины пауз в данной позиции составляет около 0,85 секунды. Мода и медиана составляют, соответственно, 0,2 и 0,55 секунды. Таким образом, видно, что в приведенных примерах (13) и (14) длина пауз не превышает «нормальную» длину.

Среди пауз на границе клауз, кроме границ после конвербов на *te/de* и *-i* отдельный интерес представляют случаи с цитационными союзами и конструкциями. Вот несколько таких примеров:

(15) SHU T-1

4. *Nani o ka-oo ka?*
что ACC купить-HOR Q

5. **to** **..(0.2)** **kangae-te** **i-mas(0.3)-u.**
 QUOT *размышлять-CNV* *AUX.PRG-ADR-PRS*
 «Что купить?» - размышляет.
 (16) WAK T-1
25. **....(1.9)** **Sorenara** **minikaa** **o** **ka-oo**
 тогда *игрушечная.машина* ACC *купить-HOR*
26. **..(0.2)** **to** **omot-te,**
 QUOT *думать-CNV*
 Тогда игрушечную машину купить, подумав...
 (17) KEN T-1
15. **oshie-te** **ager-u(0.2)**
объяснять-CNV *дать-PRS*
16. **to** **i-imash-ita.**
 QUOT *сказать-ADR-PST*
 ...объясним», - сказали.
 (18) WAK T-1
18. **sou** **dat-ta**
 так *COP-PST*
19. **..(0.3)** **to** **omot-te,**
 QUOT *думать-CNV*
 Мужчина: «Так было», - подумав...

В корпусе, как видно из таблицы 1, всего 36 вхождений цитационных конструкций, поэтому невозможно сделать точных выводов о вероятности пауз перед ними. Тем не менее, из 36 случаев употребления цитационных конструкций лишь в 8, что составляет 22%, возникли паузы. Это самый низкий процент пауз для границы клауз. Заполненных пауз в данной позиции не встретилось. Из приведенных примеров только в (14) есть небольшая пауза перед цитационным союзом.

Интересен пример (15), где пауза возникает не перед цитационным союзом, а непосредственно после него. Видимо, говорящий уже решил использовать цитационную конструкцию, но задумался при выборе конкретного глагола мышления/говoreния. Как видно по остальным примерам, такие паузы также очень редки (их всего 2 в корпусе) и обычно глагол следует непосредственно за *to* и составляет с союзом одну просодическую единицу. Кроме этого примера, в корпусе есть только один случай, когда *to* отрывается от глагола:

- (19) KEN T-1
11. **..(0.4)** **to(0.3)** **...(0.8)** **ojisan** **wa** **tazune-Ø** ==
 QUOT *дядя TOP спрашивать-CNV*
 спрашивает дядя...

В (19) примере сразу несколько факторов могли вынудить говорящего сделать паузу после *to*. Во-первых, говорящий тянет сам цитационный союз. Во-вторых, предложение оборвано. Это доказывает, что у говорящего были трудности с порождением данного фрагмента, и это могло повлечь за собой паузу после *to*. Кроме того, это единственный в корпусе пример, когда цитационный союз оторван от «своего» глагола именной группой (подлежащим), что также могло повлиять на тесноту связи между ними.

Средняя продолжительность пауз перед цитационными конструкциями составляет 0,375 секунды, медиана – 0,35, а мода 0,2 секунды.

Итак, в разных синтаксических позициях вероятность появления пауз и значения, связанные с их длиной, различны. На материале рассмотренных текстов получились такие результаты:

1. На границах предложений паузы возникают в 94% случаев, 23% из них – заполненные, их средняя продолжительность составляет 1,55 секунды, медиана их значений – 1,4 секунды и мода – 1,4 секунды.
2. На границе клауз после формы на *te/de* паузы возникают в 55% случаев, 15% - заполненные, среднее значение их длины составляет 0,5 секунды, медиана их длин – 0,6 и мода – 0,2 секунды.
3. На границе клауз после формы на *-i* паузы возникают в 67% случаев, 25% - заполненные, их средняя длина равна 0,6 секунды, медиана длин – так же 0,6 секунда, а мода – 0,4 секунды.
4. На других «стопроцентных» границах клауз паузы возникают в 63% случаев, в 18% случаев они заполненные, их средняя продолжительность составляет 0,85 секунды, медиана их длин равна 0,55, а мода – 0,2 секунды.
5. На границах клауз перед цитационным союзом *to* и другими цитационными конструкциями паузы возникнут в 22% случаев, процент заполненных пауз равен нулю. Их средняя длина составляет 0,375 секунды, медиана их длин – 0,35 и мода – 0,2 секунды.

Таким образом, видно, что паузы на границе предложений возникают чаще всего и имеют максимальную продолжительность. Это показывает, что, когда говорящий заканчивает текущую иллокуцию, ему требуется

Паузация в японском языке на границах синтаксических единиц разного уровня

некоторое время на то, чтобы обдумать продолжение и начать новое предложение. Конечно же, эта пауза используется так же и для дыхания. Эти данные согласуются с данными существующих исследований по японскому языку (таких, как [Sakura, Fuji 2006]) и с результатами корпусных исследований по другим языкам (ср. [Chafe 1994], [Кривнова, Чардин 1999], [Кибрик, Подлеская 2008], [Кибрик, Подлеская 2003]).

На границах клауз, не являющихся границами предложений, паузы также достаточно частотны и возникают более, чем в половине случаев. В среднем их длина почти в два раза меньше длины пауз на границе предложений, медиана и мода их длин значительно меньше, чем у пауз на границе предложений. Это подтверждает предположение о том, что, как правило, планирование текста в большей степени происходит на границах больших синтаксических единиц, а паузы на границах клауз, не равных границам предложений, могут использоваться говорящим для дыхания, обозначения синтаксических границ и, в меньшей степени, для обдумывания продолжения своего текста. Этот результат также совпадает с результатами предыдущих работ в этой области (см. [Sakura, Fuji 2006]).

Что касается цитационных конструкций, то вероятность и длина пауз на границах клауз перед ними минимальна. Видимо, планируя предложение с цитатой, говорящий уже знает, что это цитату надо будет оформить специальным образом, и ему не требуется дополнительное обдумывание на границе клауз. На это указывает также отсутствие заполненных пауз перед цитационными конструкциями. Цитационные конструкции ранее отдельно не рассматривались на предмет возможности пауз до и внутри них, и в этой области данная работа предоставляет новые данные.

Список литературы

1. Кибрик А.А., Подлеская В.И., К созданию корпусов устной русской речи: принципы транскрибирования // Научно-техническая информация. Серия 2: Информационные процессы и системы, № 10. Москва, 2003, 5-11.
2. Кибрик А.А., Подлеская В.И. (ред.) Рассказы о сноведениях: корпусное исследование устного русского дискурса. Москва, 2008. (в печати)
3. Кривнова О.Ф., Чардин И.С. Паузирование при автоматическом синтезе речи // Теория и практика речевых исследований (АРСО-99). Москва, 1999.
4. Хуршудян В.Г. Средства выражения гезитации в устном армянском дискурсе в типологической перспективе. Диссертация на соискание ученой степени кандидата филологических наук. Москва, 2006.
5. Chafe. W. Discourse, consciousness, and time. Chicago, 1994.
6. Iwasaki S. The Structure of the Intonation Unit in Japanese // Japanese/Korean Linguistics, vol. 3, ed. by Soonja Choi. Stanford, 1993, 39-53
7. Sakura C., Fuji S. Intonation units, information structure, and grammatical constructions in Japanese and English. // Oral presentations. University of Tokyo, 2006.