

ОПЫТ ПРИМЕНЕНИЯ МЕТОДОВ ДИСКУРСИВНО ОРИЕНТИРОВАННОЙ ТРАНСКРИПЦИИ К МАТЕРИАЛУ НЕИНДОЕВРОПЕЙСКОГО ЯЗЫКА¹

APPLYING METHODS OF THE DISCOURSE-ORIENTED TRANSCRIPTION TO THE NON-INDO-EUROPEAN LANGUAGE

В. Л. Цуканова (zukanova@mail.ru)
РГГУ

В докладе рассматриваются особенности применения дискурсивной транскрипции, разработанной для корпуса русских текстов, к текстам на кувейтском диалекте арабского языка. Основное внимание уделяется нестандартному делению на дискурсивные единицы и связанным с этим грамматическим особенностям.

Доклад посвящен применению методов транскрипции, разработанной для анализа корпуса устной детской речи (рассказы о сновидениях) к материалу корпуса устных рассказов на кувейтском диалекте арабского языка.

Разработка транскрипции для русских устных рассказов ведется с 2003 г. исследовательской группой под руководством А.А. Кибрика и В.И. Подлесской. Разработанные принципы транскрибирования уже применялись к материалу других языков, например, к языку пулар [Коваль&Кибрик 2005], [Кибрик и др. 2006]. Дискурсивная транскрипция корпуса нарративов на кувейтском диалекте арабского языка осуществляется в рамках диссертационной работы «Употребление показателей структуры дискурса в устных нарративных текстах (на материале кувейтского диалекта арабского языка)». Она опирается на уже готовые критерии, разработанные для русского языка, однако при применении их естественным образом возникает ряд трудностей ввиду несовпадения синтаксиса и интонационной структуры двух языков.

В дискурсивно ориентированной транскрипции отмечаются прежде всего параметры, релевантные для дискурсивного анализа и построения ТРС-графов². Одной из проблем при разработке такой транскрипции является деление текста на элементарные дискурсивные единицы (далее – ЭДЕ). Дискурсивные единицы – это кванты, которыми порождается дискурс и которые являются узлами ТРС-графов. В дискурсивной транскрипции каждая ЭДЕ записывается отдельной строкой. Существует два критерия выделения ЭДЕ: семантико-синтаксический и интонационный. В соответствии с первым критерием ЭДЕ должна соответствовать предикации, в соответствии со вторым ЭДЕ должна совпадать с интонационной единицей. Основные проблемы при разбиении на ЭДЕ возникают в случае несовпадения этих двух критериев. Для русского языка выделен ряд типовых случаев, когда ЭДЕ меньше или больше предикации. Принципы транскрибирования довольно подробно рассмотрены в [Кибрик, Подлесская 2003].

Другая проблема связана с представлением в дискурсивной транскрипции текста на иностранном языке, требующим глоссирования. Одна ЭДЕ должна быть, таким образом, записана четырьмя строками: первая представляет собственно дискурсивную транскрипцию, в которой отмечаются паузы, акценты, хезитации, фальстарты, пунктуационные знаки и некоторые другие явления. Второй строкой идет поморфемное представление, третьей глоссы и четвертой перевод:

(1) S2
 ee(0.6) ..(0.2) ›ihna kinna kill /ʊɟm ee(0.6) nidfa◁ \°adaqah.
 ›ihna kin-na kill yɟm ni-dfa◁ °adaq-ah
 мы быть.PFV-1PL каждый день.SG 1PL-платить.IPFV милостыня-F.SG

Мы каждый день платили милостыню.

Ввиду того, что проблематикой доклада является деление на ЭДЕ, мы не будем приводить первую строку полностью, а будем лишь отмечать в строке, содержащей поморфемное членение, релевантные для рассматриваемых вопросов параметры. Если пример содержит несколько ЭДЕ, то перевод приводится отдельно для каждой из них. Перед каждым примером стоит номер текста, принятый в корпусе.

¹ Исследование поддержано грантом РГНФ № 04-04-00220А.

² ТРС-граф – граф, отражающий дискурсивные отношения между частями текста в терминах Теории риторических структур Манна и Томпсон [Mann, Thompson 1988].

ЭДЕ, совпадающие с предикациями

Приведем по одному примеру наиболее типичного случая из обоих корпусов, где ЭДЕ совпадают с предикациями, и выделение их не вызывает сомнений:

(2) N42

/Ну и я шла по этому /саду,
 ..(0.4) и /куда-то дошла до какой-то \двери.
 ...(0.9) Поднялась по /лестнице,
 и \дошла вот до этой \двери.

(3) S1

jna gjkid ʔa-dris,
 я PROG 1SG-учиться.IPFV
 «[...] я учусь,

ʔjna ʔa-ħibb id-dirjs-a!
 я 1SG-любить.IPFV DEF-учеба-F.SG
 я люблю учебу!

Uu ʔenti sawwi-t-i l-baħt hj-ɖa,
 и ты.F делать.II.PFV-2-SG DEF-задание.SG этот-M.SG
 А ты сделала это задание,

illī gjl-ɖ l#es ʔann#aʔʔ
 который сказать.PFV-3PLʔ#тебе.F о#нем
 о котором тебе сказали?»

ЭДЕ, меньшие, чем предикации

В русском корпусе к ним относятся: вынесенные топики, парцелляции (вербализация составляющей после основной части предикации), разрыв предикации, фальстарт (говорящий отказывается от ЭДЕ, которую он уже начал строить), регуляторные ЭДЕ (состоящие исключительно из дискурсивных маркеров).

Топикализация

Вынесенный топик в русском корпусе считается самостоятельной ДЕ, если он представляет собой вынесенную в начало клаузы именную группу (обычно в именительном падеже), несущую восходящий акцент, которая предшествует основной предикации и имеет в ней кореферентную местоименную копию:

(4) Z33

..(0.3) /Зайчик вот,
 ...(0.5) он был в \лесу.

В случае, когда такая ИГ не несет на себе акцента, или имеет падающий тон, или не имеет кореферентной копии внутри последующей клаузы, она не рассматривается как вынесенный топик и в отдельную строку не выделяется.

В кувейтском корпусе наблюдается несколько иная ситуация. Топик может предшествовать или следовать за именной группой, имеющей кореферентную ему местоименную копию, причем он может быть вынесен не только в начало или конец всей предикации, но и в середину ее.

Вынос топика в конец клаузы, на топике восходящий акцент:

(5) S1

\ħallam-t#nī,
 присниться.II.PFV-1SG#мне
 приснилась я себе

/ʔjna,
 я
 я

Вынос топика перед ИГ с кореферентной местоименной копией, на топике восходящий акцент:

(6) S1

Ya-ɖnī ma-frʔ!

3.М-значить.IPFV[SG] РТСР-предполагать.P.SG
Значит, предполагается

sjh-a-t /il-madras-a,
площадь-F.SG-C DEF-школа-F.SG
что площадь школы,

ʔa-lʔf#hj \riddjm#i,
1SG-смотреть.IPFV#ee перед#мной
увидю ее перед собой,

Вынос топика перед ИГ с кореферентной местоименной копией внутри одной предикации, на топике нет акцента:

(7) P2
/Rjh -ye-lʔf,
пойти.PFV[3SG.M] 3.М-увидеть.IPFV[SG]
Пошел посмотреть,

lʔnʔ zɔr-t#a \jinda#ha.
что жена-F.C#его у#нее
что есть у его жены.

Представляется осмысленным для кувейтского диалекта выделять вынесенный топик в отдельную ЭДЕ только в том случае, если он вынесен в начало или конец предикации и на нем есть акцент, то есть он представляет собой независимую интонационную единицу, вне зависимости от его позиции по отношению к ИГ с местоименной копией.

Регуляторные ЭДЕ

Регуляторные ЭДЕ состоят из дискурсивных маркеров, регулирующих структуру дискурса. Они делятся на несколько классов, из них в отношении кувейтского корпуса интересны финальные и подготовительные.

Финальные ДЕ сигнализируют о завершении некоторого фрагмента. В примере (8) из русского корпуса эту роль выполняет *вот* в сочетании с другим дискурсивным маркером:

(8) Z12
...(0.7) /Ну я там /помню,
..(0.1) что-о' я там(1.1) ну с моим /другом ...(0.5) \разговаривал я.
...(2.9) Ну –вот.
/Потом ==

В кувейтском диалекте ту же роль выполняет маркер *nzeyn*:

(9) S1
ul#ʔa-lʔf sjh-a-t il-malʔab,
и#1SG-смотреть.IPFV площадь-F.SG-CDEF-стадион.SG
а я вижу площадь стадиона,

/nzɪn,
NZEYN
вот,

u#jna balj gʔti,
и#я без туфли.PL
и я без туфель,

Похожую функцию выполняет маркер *ilmuhimm*, стоящий в начале эпизода. Оба маркера могут как нести на себе акцент, так и составлять одну интонационную единицу с главной предикацией. Тем не менее, в обоих случаях они выделяются в отдельную ЭДЕ.

(10) S2
\l-muhimm,
DEF-главное.SG
Ну вот,

\waddey-t zɔr#i ==

отвезти.П.PFV-1SG муж.SG#мой
я отвезла моего мужа

raddey-t#a l-beyt,
вернуть.П.PFV-1SG#его DEF-дом.SG
отправила его домой

То же в случае сочетания этих маркеров с другими маркерами или служебными словами:

(11) P5
Bass >inn#a /l-muhimm,
но что#его DEF-главное.SG
Но что главное,

љагј /s-sayujr-a,
купить.PFV[3SG.M] DEF-машина-F.SG
он купил машину,

Подготовительные дискурсивные маркеры (препаративная замена) используются говорящим при поиске вербализации, если говорящему уже известна часть речи или синтаксическая структура искомого фрагмента. В русском корпусе встречаются изменяемые маркеры *этой, такой* и неизменяемые *это* и *так*. Эти же типы мы видим и в кувейтских текстах: *lišma* 'как его' и *hǰda* 'этот' могут согласовываться с искомым именем в роде и числе, маркер *sedī* 'так' неизменяем.

В примере (12) *hǰda* находится в середине ЭДЕ и согласуется в роде с существительным:

(12) S2
°j'kid-a <ala s-sayujr-a li-zfīr-a \hǰ-dī
подняться.ПТСР.A-F.SG на DEF-машина-F.SG DEF-маленький-F.SG этот-F.SG
...(0.5) ee(0.4) m= mj-l-t el-yahhjl,
ATTR-F.C DEF-ребенок.PL
поднялась на маленькую машину эту... детскую,

ср. в русском корпусе:

(13) N54
..(0.3) «/Алёна \Игоревна@
А /вы –сегодня нас /поведёте в \этот ..(0.4) в –тренажёрный \зал?».

lišma в середине ЭДЕ, не согласуется с последующим существительным:

(14) P4
Ba:di:n sjn \yu-waddi#ha љ# \ism#a hagg /zur-t#a,
потом AUX 3М-нести.П.IPFV[SG]#ее что#имя.SG#его для жена-F.C#его
Потом отнес ее это... к своей жене,

В кувейтском корпусе ни разу не встречается маркер препаративной замены в начале клаузы, и в отдельную регуляторную ЭДЕ они не выделяются. Для сравнения: в русском корпусе в отдельную ЭДЕ выделяются ударные неизменяемые подстановки *это* и *так*, если они стоят в начале клаузы:

(15) Z56
..(0.3) \Это,
..(0.2) /бизнес-класс ..(0.4) и' ...(0.5) и' \дорогие классы.

Эти же маркеры могут встречаться в конечной позиции в мужском роде в значении незаконченного перечисления:

(16) P11
mj gam ya-<ni <yi>-meuyiz,
NEG статья.PFV[3SG.M] 3М-значить.IPFV[SG] 3М-различать.П.IPFV[SG]
Перестал, значит, различать

љinu riddjm#<a> u#/hǰ-da,
что перед#ним и#этот-M.SG
что перед ним и все такое,

(17) S1
u=fī=hǰ mdarrar-jt u=sedī _

и=v=ней ступенька-F.PL и=так
и на ней трибуны и прочее

В этой позиции они не выделяются в отдельную ЭДЕ.

Универсальными для обоих языков оказываются проблемы парцелляций и фальстартов.

Парцелляции представляют собой случаи расщепления дискурсивной единицы на две, когда к уже завершённой предикации добавляется интонационно выделенный фрагмент.

Постпозитивные определения в русском языке практически всегда рассматриваются как парцеллированные:

(18) N47

...(0.6) Я' ..(0.4) \зашёл в /комнату,
в \большую,

В кувейтском диалекте их реже можно интерпретировать таким образом, так как прилагательные и другие виды определений (изафетная конструкция, бессоюзные определительные придаточные) следуют обычно в постпозиции к имени, так что разбиение на две ЭДЕ возможно только при очевидном интонационном выделении:

(19) P5

<arb-at#a \sayyjг-a,
нравится.PFV-3SG.F#ему DEF-машина-F.SG
понравилась ему машина,

....(1.4) \hilw-a,
красивый-F.SG
красивая,

lцn#ha \hilu,
цвет.SG#ee красивый.SG
красивого цвета,

..(0.3) u#лакil#ha \hilu,
и#форма.SG#ee красивый.SG
красивой формы,

u#\spггт.
и#спорт
и спортивная.

Остальные виды парцелляций – дополнений, обстоятельств, сочиненных групп – встречаются в обоих корпусах, ср., например, парцелляцию сочиненных именных групп:

(20) N32

...(0.5) "(0.2) ... (0.9) И /он с 'эххэ(0.6) с-с ..(0.4) \длинными ... (0.6) "'(0.6) такими выющи= ээ(0.2) =мися
"(0.3) \светлыми /волосами,
...(0.7) ээ(0.2) и с \сумкой ээ(0.1) за \плечами.
...(0.5) {ЦОКАНЬЕ} ..(0.2) И || и с \верёвкой' в \руках.

(21) P3

Ligj <inda#ha \ьенj© u#add-jt \makiyjнь u#kull \ьеу.
нашел.PFV[3SG.M] у#нее сумка.PL и#предмет-F.PL макияж.SG и#все
вещь.SG
Нашел у нее сумки, и предметы макияжа, и всё.

...(0.5) U#-sa<-jt u#\zjin-a.
и#часы-F.PL и#украшение-F.SG
И часы, и украшения.

Фальстарты, по-видимому, представляют собой универсальное языковое явление, которое мы здесь подробно рассматривать не будем. Приведем пример фальстарта, влияющего на деление ЭДЕ:

(22) S3

°гг еl= ==
статья.PFV[3SG.M] DEF
Настал ==

Кjn fi љufl hni uɬm ʔjʎʔga.
 быть. PFV[3SG.M] есть работа.SG здесь день.SG Ашура
 Была здесь работа в день Ашура.

ЭДЕ, большие, чем предикации

ЭДЕ большие, чем предикации, возникают в случае редукции одного из предикатов до статуса цитационного или эпистемического маркера, в некоторых сложных предложениях и при сериализации.

Цитации

В кувейтском корпусе, в отличие от русского, глагол, вводящий прямую речь, не встречается после или внутри цитации, а всегда предшествует ей. Часто он бывает неакцентирован, хотя не редуцируется так сильно, как в русском, всегда представляет собой отдельное фонетическое слово и поэтому записывается отдельной строкой. Так, в примере (23) на вводящем предикате нет акцента, несмотря на то, что он представляет собой аналитическую форму глагола, а в примере (24) есть, хотя это более короткое фонетическое слово, а формы в данном контексте синонимичны:

(23) P2
 Sjn yi-gʔl:
 AUX 3.M-сказать.IPFV[SG]
 Сказал:

“/Wɬɬid ʔjli!”
 очень дорогой.SG
 «Очень дорого!»

(24) P5
 \Gɬl:
 сказать.PFV[3SG.M]
 Сказал:

“Lj ɬj ʎa@ɬ.”
 NEG NEG всё
 «Нет, нет, всё.»

Ср. пример из русского корпуса, где цитационный маркер сильно редуцирован и не выделяется в отдельную ЭДЕ:

(25) N54
 говорю.«/Давай я всё \сделаю!»

В случае перестройки косвенной речи в прямую союз, вводящий косвенную речь, записывается одной строкой с последующей цитацией:

(26) P5
 fa ga< yi-gʔl lu#hum,
 и PROG 3.M-сказать.IPFV[SG] к#ним
 и говорит им,

enp#a “ɬjna mi-htjɬ.”
 что#он я ПТСР-быть.в.замешательстве.A.SG
 что «Я в замешательстве.»

Эпистемические маркеры

Существует ряд предикатов, которые лексикализовались в эпистемические маркеры и не выделяются в отдельную ЭДЕ. Самыми частотными из них в кувейтском корпусе являются *ya<ni* ‘значит’ и *wa@@a* ‘ей-Богу’:

(27) S2
 Wa#@@a ya<ni hj-da al-hin a-rawwi#s
 iuyj#0,
 ей#Бог.SG 3М-значить.IPFV[SG] этот-SG.M DEF-время.SG 1SG-показать.II.IPFV#тебе.F ACC#его
 Ей-Богу, значит, это сейчас покажу тебе.

Ср. пример из русского корпуса, содержащий лексикализованные предикаты:

(28) Z55
 Значит мне /снится,
 что я где-то на –приро-оде_

Both ..(0.3) кажется на /даче,
но /я вот там ни разу не \была.

Проблема сложных предложений.

В кувейтском корпусе встречаются сложные предложения такой структуры, в которой сложно определить границы клауз и, соответственно, ЭДЕ.

Прежде всего, это предложения с сочинительным союзом *и* (*ul*), зависимая часть которых часто содержит не финитный глагол, а причастие или нулевой предикат. Тем не менее, подобные предложения в арабском языке являются сложноподчиненными [Бронников 1998].

(29) P8

ħaḡam maḡ-a t̪niy-a
запретить.П.PFV[3SG.M] раз-F.SG второй-F.SG
Зарекся в другой раз

ʔənn#a /ȳ-lʔab,
что#он 3.M-играть.IPFV[SG]
заниматься (спортом),

u#ħuwa \ʔj̄ib.
и#он пить.PTCP.A.SG
когда выпьет (= и он выпивший).

(30) P2

K̪j̄n fi /rauyj̄l,
быть.PFV[3SG.M] есть мужчина.SG
Жил-был мужчина

...(0.7) u#maʔ \ʔj̄il-a
и#c семья-F.SG
с семьей.

В примере (30), помимо синтаксических, есть еще просодические критерии, по которым следует рассматривать эти части предложения как две ЭДЕ: наличие двух несущих акцентов. Однако возможна ситуация (не встретившаяся в корпусе), когда акцент на зависимой части отсутствует. В таком случае интонационный критерий будет превалировать над синтаксическим, и обе части будут объединены в одну ЭДЕ.

Другой вид сложных предложений, представляющий определенные трудности – это определительные придаточные при отсутствии относительного местоимения или с нулевым предикатом.

В примере (31) относительное местоимение невозможно ввиду того, что вершинное существительное ‘мужчина’ стоит в неопределенном состоянии; обе клаузы составляют одну интонационную единицу:

(31) P6

\k̪j̄n aku \ħnj̄k rauyj̄l ism#a /Miʔj̄ri,
быть.PFV[3SG.M] есть там мужчина.SG имя.SG#его Мишари
Там был человек по имени Мишари,

В примере (32) присутствует относительное местоимение, но в придаточном – нулевой предикат, и один несущий акцент на все предложение:

(32) P6

U#baʔdiyn be#ʔedf-a huwa g̪ʔid ȳi-ʔȳf il-ʔaʔuj
и#потом с#неожиданность-F.SG он PROG 3M-видеть.IPFV[SG] DEF-вещь.PL

illi d̪j̄j̄il /il-vitr̄ina,
который внутри DEF-витрина

А потом он неожиданно видит вещи внутри витрины,

Есть и обратный пример, с нулевым предикатом в главной клаузе:

(33) S2

Is= iʔnaʔaʔal bi#l-liyl illi \ʔkey ʔjr.
двенадцать в#DEF-ночь.SG который окей стать.PFV[3SG.M]
В двенадцать ночи наладилось.

Во всех вышеприведенных примерах сложноподчиненные предложения записываются как одна ЭДЕ, что, как нам кажется, соответствует психологической реальности. Однако есть примеры относительных придаточных,

представляющих разные интонационные единицы и записывающиеся в две ЭДЕ. Так, в примере (34) в главной части нулевой предикат, а в зависимой – нефинитная глагольная форма (страдательное причастие), но есть два четко выраженных несущих акцента:

(34) P3
 ʎinʝ /l-ʌʎuʝ,
 что DEF-вещь.PL
 что за вещи

 illi /ma-wrʝd-a \<inda#ha.
 который PTCP-находить.P-F.SG у#нее
 имеются у нее.

Следующий тип связи между предикатами, не представляющий особой трудности, но на который необходимо обратить внимание – это бессоюзное подчинение финитных форм глагола в сентенциальных актантах:

(35) P1
 rʝh /уе-ʎʝf >aʝtʝi /zʝr-t#a,
 пойти.PFV[3SG.M] 3.M-увидеть.IPFV[SG] вещь.PL жена-F.C#его
 пошел посмотреть вещи своей жены,

(36) S3
 mj gjm \yi-ʎʝf.
 NEG стать.PFV[3SG.M] 3.M-увидеть.IPFV[SG]
 перестал видеть.

Такие случаи всегда записывают как одна ЭДЕ. В русском языке им обычно соответствуют инфинитивные конструкции:

(37) N48
 Он стал носить эту /лягушку на \голове.

Также общим явлением для двух языков оказывается сериализация предикатов, ср.:

(38) P3
 Fa#rʝh ʎʝf fi#ʝurf-a-t#ha
 и#пойти.PFV[3SG.M] увидеть.PFV[3SG.M] в#комната-F.SG-C#ее
 И пошел посмотрел в ее комнату

(39) N54
 мы /взяли \поменяли /кровати,

Заключение

Принципы дискурсивно ориентированной транскрипции, разработанные для русского языка, в основном успешно применяются к материалу языков другого типа и позволяют выявить некоторые их дискурсивно релевантные особенности.

По умолчанию (не отмечаются ввиду наибольшей частотности): у имен - мужской род, неопределенное состояние, у глаголов: I порога, активный залог.

Условные обозначения

#	граница клитики	AUX	вспомогательный глагол
=	фальстарт	C	сопряженное состояние
/, \, ʌ	направление движения тона	DEF	определенный артикль
..(0.1)	паузы	F	женский род
@	обращение	IPFV	имперфектив
i	директив	M	мужской род
a	несущий акцент	NEG	отрицание
1,2,3	1-е, 2-е, 3-е лицо	P	пассив
II, III... X	производные порога глагола	PFV	перфектив
A	актив	PL	множественное число
ACC	аккузатив	PROG	прогрессив
ATTR	атрибутив	PTCP	причастие

Список литературы

1. *Бронников Г.К.* Предложения с союзом wa в арабском языке. Дипломная работа. М.: РГГУ, 1998.
2. *Кибрик А.А., Подлесская В.И.* К созданию корпусов устной русской речи: принципы транскрибирования. // Научно-техническая информация, серия 2, №10, 2003 г., стр 5-12.
3. *Кибрик А.А., Ба А.Й., Коваль А.И.* Сказание «любовь», или «дингирал», на языке пулар. // Африканская сказка – III. К исследованию языка фольклора. М.: Восточная литература, 2006 (в печати). С. 399-479.
4. *Коваль А.И., Кибрик А.А.* Дискурс гриота. // Четвертая типологическая школа. Тезисы докладов. М, 2005.
5. *Mann W.C., Thompson S.A.* 1988. Rhetorical structure theory: toward a functional theory of text organization. Text, 8, 243-281.