



**HAL**  
open science

**Проблемы транскрибирования устных армянских текстов  
для корпусных исследований [Problèmes de transcription  
des textes oraux arméniens pour les recherches de corpus]**

Victoria Khurshudyan

► **To cite this version:**

Victoria Khurshudyan. Проблемы транскрибирования устных армянских текстов для корпусных исследований [Problèmes de transcription des textes oraux arméniens pour les recherches de corpus]. NTL, 2007, 2, pp.37-46. <halshs-00722504>

**HAL Id: halshs-00722504**

**<https://shs.hal.science/halshs-00722504>**

Submitted on 2 Aug 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

В. Г. Хуршудян

## **Проблемы транскрибирования устных армянских текстов для корпусных исследований<sup>1</sup>**

*Работа посвящена проблемам письменной фиксации устных армянских текстов для нужд корпусных исследований. Детально рассматриваются два собранных автором небольших электронных корпуса устных рассказов и диалогов на армянском языке, на материале которых была отработана предлагаемая система разметки и нотации, в том числе, основные принципы транслитерации и глоссирования.*

Задача данной работы – предложить ряд специальных решений, которые позволили бы компактно и адекватно представлять устные тексты на армянском языке в транскриптах, или, стенограммах, в составе электронных корпусов. Более точно, объектом нашего исследования является современный восточно-армянский язык (язык основной части населения современной Армении). Данная работа выполняется в рамках проекта по созданию универсального стандарта дискурсивной транскрипции для многоязычных корпусов [1, 2, 3].

В дискурсивном анализе имеется ряд методов транскрибирования для разных жанров устного дискурса, см. [4, 5, 6, 7, 8], более подробный обзор методов транскрибирования см. [9,10]. Поскольку существующие методы отработаны на материале очень ограниченного числа языков, их применение к материалу более редких языков вызывает необходимость внести некоторые изменения и дополнения в формат транскрипции.

Основным материалом для данного исследования послужила армянская часть собранного нами четырехязычного корпуса устных нарративов на армянском, русском, итальянском и английском языках (далее АКК – Армянский корпус по картинкам). Для сбора текстов были выбраны два рассказа по картинкам. Участвовали 10 информантов по каждому языку-объекту. Возраст информантов в среднем 20-30 лет. Записи проводились в 2003-2004 гг. в Ереване и в Москве (носители итальянского и английского языков также были записаны в Ереване и в Москве). Задание состояло из двух этапов. В самом начале информанту сообщалось, что эксперимент состоит из двух частей, и что вторую часть надо продолжить через определенный промежуток времени (6-8 часов), суть второй части эксперимента не объяснялась. Сначала информантам давалось несколько секунд на просмотривание одной серии картинок, чтобы у них возникло общее представление о сюжете, после чего информанты составляли рассказы по этим картинкам, имея их перед глазами. После 6-8 часов информантов просили пересказать те же рассказы, но, уже не видя картинки, по памяти. Аналогичным образом проводился эксперимент и по второй серии картинок. Корпус состоит из 160 текстов, с общим объемом приблизительно 20.000 словоупотреблений и представлен двумя основными частями: аудиозаписи и их транскрипты.

Нами дополнительно был собран корпус спонтанных бытовых диалогов на армянском языке, который выступает в качестве вспомогательного корпуса (далее АКД – Армянский корпус диалогов). Этот корпус записан нами в Ереване в январе

---

<sup>1</sup> Исследование выполнено при поддержке РФФИ (грант № 06-06-80470а)

2006 г. Участвовало 5 носителей армянского языка, возраст информантов 20-30 лет. Общий объем корпуса насчитывает приблизительно 20.000 словоупотреблений. Данный корпус анализировался менее детально и его материал преимущественно использовался при описании сегментных средств хезитации [11].

Относительно имеющихся работ и традиций исследования по устному дискурсу в армянском языке, стоит отметить, что на эту тему имеется всего несколько публикаций, которые, как правило, имеют дескриптивный характер и рассматривают наиболее общие вопросы [12, 13]. На самом начальном этапе находится и работа по созданию электронных корпусов армянских текстов<sup>2</sup>, поэтому практически отсутствуют и корпусные исследования, и, следовательно, нет определенных стандартов транскрибирования, которые отвечали бы этим задачам. В настоящей работе мы пытаемся частично восполнить этот пробел и сформулировать основные принципы, касающиеся следующих аспектов транскрипции:

1. *Графическое представление армянских текстов,*
2. *Грамматическая аннотация, перевод,*
3. *Сегментация устных нарративов и общие принципы транскрибирования.*

Рассмотрим пути решения этих вопросов более подробно.

## **1. Графическое представление армянских текстов**

Поскольку армянский язык имеет оригинальную графику, необходимо было представление текстов в транслитерации, параллельно с армянской графикой. Такой способ представления армянских текстов делает их доступными для более широкого круга специалистов. Оригинальная графика представлена в юникодной кодировке Sylfaen (о графической системе армянского языка более подробно см. [14]).

Для передачи армянского алфавита на латинице (имеются системы также на кириллице) в арменистике были созданы несколько систем транслитерации, см. Таблицу 1. Классической и стандартной системой транслитерации армянского языка считается система, предложенная Гюбшманном и Мейе [15], на основе которой и выработана система транслитерации, примененная в нашей работе. Поскольку система Гюбшмана-Мейе была создана для древнеармянского языка, ее применение к современному армянскому языку потребовало частичных изменений. В этих изменениях мы основывались на принципах, предложенных Международной фонетической ассоциацией [16], что позволило более точно отразить фонологические особенности армянского языка. В частности:

- в нашей системе придыхание отображается надстрочной графемой <sup>h</sup>, например *‘t<sup>h</sup>’ - /t<sup>h</sup>/, ‘p<sup>h</sup>’ - /p<sup>h</sup>/* и т.д.
- для отображения графем *‘ղ’*, *‘խ’* употребляется *‘v’*, *‘x’* соответственно, что показывает, что они являются увулярными согласными (а не велярными). В

---

<sup>2</sup> В настоящее время под руководством д.ф.н. В.А. Плунгяна (Институт языкознания РАН) ведутся работы по созданию Восточно-армянского национального корпуса, который будет иметь подкорпус устных текстов. Наличие подобного корпуса, безусловно, будет существенным инструментом изучения как устной, так и письменной армянской речи. Также по инициативе Анаид Донабедян (INALCO, Париж) был создан небольшой корпус (около 1,5 млн. словоупотреблений) письменных текстов западно-армянского языка для лингвистических исследований (<http://a.donabedian.free.fr/textes/Noojarmenien.pdf>). Кроме того, имеются некоторые корпуса древнеармянских письменных текстов, которые, как правило, представляют собой только банк текстов ([www.digilib.am](http://www.digilib.am), <http://titus.uni-frankfurt.de>, <http://www.indo-europees.leidenuniv.nl/index.php3?c=64>).

системе Гюбшмана-Мейе употребляются графемы /h/ и /x/ соответственно. Такое представление объясняется тем, что в древнеармянском /h/ имел качество бокового велярного сонанта, который в современном армянском превратился в увулярную фрикативную согласную.

- Отличием от системы Гюбшмана-Мейе представляется транслитерация аффрикат ‘d’ – ‘j’ и ‘g’ – ‘j’<sup>3</sup>, которые, согласно МФА [16], в нашей системе передаются графемами ‘z’ и ‘z’<sup>3</sup> соответственно.
- Поскольку в фонологической системе древнеармянского языка не было монофтонга /u/, а также соответствующей графемы, в системе транслитерации Гюбшмана-Мейе отсутствует отдельное обозначение для данной фонемы. В современном армянском языке имеется простая гласная /u/, которая передается диграфом ‘ni’. Соответственно, в нашей системе диграф ‘ni’ транслитерируется знаком ‘u’.
- Ввиду того, что в нашей работе транслитерация выполняла практические задачи, мы учли особенности некоторых букв, которые имеют разное прочтение в зависимости от позиции, в частности это касается букв ‘k’, ‘n’, ‘l’, которые могут быть представлены, как ‘ye/e’, ‘vo/o’, ‘yev/ev’ соответственно.

Ниже, в Таблице 1 представлена общая система транслитерации, примененная в нашей работе. В конце статьи, в Таблице 3 приводится образец затранскрибированного текста, в котором продемонстрировано общее применение данной системы, а также распределение позиционно-вариативных графических значений.

Таблица 1. Системы транслитерации для армянского языка

| Армянская графика |      | Наша система транслитерации |                | Другие системы транслитерации |                 |                       |                  |                   |
|-------------------|------|-----------------------------|----------------|-------------------------------|-----------------|-----------------------|------------------|-------------------|
| Загл.             | Стр. | Загл.                       | Стр.           | Г-М <sup>3</sup>              | LC <sup>4</sup> | BGN/PCGN <sup>5</sup> | ISO <sup>6</sup> | KNAB <sup>7</sup> |
| Ա                 | ա    | A                           | a              | a                             | a               | a                     | a                | a                 |
| Բ                 | բ    | B                           | b              | b                             | b (p)           | b                     | b                | b                 |
| Գ                 | գ    | G                           | g              | g                             | g (k)           | g                     | g                | g                 |
| Դ                 | դ    | D                           | d              | d                             | d (t)           | d                     | d                | d                 |
| Ե                 | ե    | Ye, E                       | ye, e          | e                             | e, y            | e, ye                 | e                | e, ye             |
| Զ                 | զ    | Z                           | z              | z                             | z               | z                     | z                | z                 |
| Է                 | է    | Ē                           | ē              | ē                             | ē               | e                     | ë                | è                 |
| Ը                 | ը    | Ə                           | ə              | ə                             | ě               | y                     | ə                | y                 |
| Թ                 | թ    | T <sup>h</sup>              | t <sup>h</sup> | t’                            | t’              | t’                    | t’               | th                |
| Ճ                 | ճ    | Ž                           | ž              | ž                             | zh              | zh                    | ž                | ž                 |
| Ի                 | ի    | I                           | i              | i                             | i               | i                     | i                | i                 |
| Լ                 | լ    | L                           | l              | l                             | l               | l                     | l                | l                 |
| Խ                 | խ    | χ                           | χ              | x                             | kh              | kh                    | x                | x                 |

<sup>3</sup> Система транслитерации, предложенная Гюбшманном и Мейе [15]

<sup>4</sup> Система транслитерации American Library Association – Library of Congress, 1997 г.

<sup>5</sup> Система транслитерации United States Board on Geographic Names, 1981 г.

<sup>6</sup> Система транслитерации International Standards Organization ([www.iso.ch](http://www.iso.ch)), 1996 г.

<sup>7</sup> Система транслитерации Place Names Database of Institute of the Estonian Language, 1993 г.

|         |        |                |                |    |         |         |    |    |
|---------|--------|----------------|----------------|----|---------|---------|----|----|
| Ծ       | ծ      | С              | c              | c  | ts (dz) | ts      | ç  | ts |
| Կ       | կ      | К              | k              | k  | k (g)   | k       | k  | k  |
| Հ       | հ      | Н              | h              | h  | h       | h       | h  | h  |
| Ձ       | ձ      | Յ              | ɟ              | j  | dz (ts) | dz      | j  | dz |
| Ղ       | ղ      | Ե              | ɛ              | ɫ  | gh      | gh      | ğ  | ğ  |
| Ճ       | ճ      | Շ              | č              | č  | ch (j)  | ch      | č  | č  |
| Մ       | մ      | М              | m              | m  | m       | m       | m  | m  |
| ԅ       | յ      | Y              | y              | y  | y, h    | y       | y  | j  |
| Ն       | ն      | Н              | n              | n  | n       | n       | n  | n  |
| Շ       | շ      | Տ              | š              | š  | sh      | sh      | š  | š  |
| Ո       | ո      | Vo, O          | vo, o          | o  | o       | o, vo   | o  | o  |
| Չ       | չ      | Č <sup>h</sup> | č <sup>h</sup> | č' | ch'     | ch'     | č  | čh |
| Պ       | պ      | Р              | p              | p  | p (b)   | p       | p  | p  |
| Ջ       | ջ      | Ջ              | ǰ              | ǰ  | j (ch)  | j       | ǰ  | dž |
| Ռ       | ր      | Ř              | ř              | ř  | ř       | rr      | ř  | ř  |
| Ս       | ս      | S              | s              | s  | s       | s       | s  | s  |
| Վ       | վ      | V              | v              | v  | v       | v       | v  | v  |
| Տ       | տ      | T              | t              | t  | t (d)   | t       | t  | t  |
| Ր       | ր      | R              | r              | r  | r       | r       | r  | r  |
| Յ       | յ      | C <sup>h</sup> | c <sup>h</sup> | c' | ts      | ts      | c' | ch |
| Ու      | ու     | U              | u              | w  | w       | -       | w  | w  |
| Փ       | փ      | P <sup>h</sup> | p <sup>h</sup> | p' | p'      | p'      | p' | ph |
| Ք       | ք      | K <sup>h</sup> | k <sup>h</sup> | k' | k'      | k'      | k' | kh |
| Եվ (Եւ) | և (ևւ) | Yev            | yev, ev        | ev | ew      | ev, yev | ew | ew |
| Օ       | օ      | Ō              | ō              | ō  | ō       | o       | ò  | ò  |
| Ֆ       | ֆ      | f              | f              | f  | f       | f       | f  | f  |

## 2. Грамматическая аннотация, перевод

В тексте каждой словоформе приписывается ее лемматизированный перевод и соответствующая грамматическая аннотация. Выделяются все значимые грамматические категории, которые размечаются определенными тэгами (в целом, в соответствии с Лейпциговской системой глоссирования [17]). В качестве рабочего решения было принято рассматривать признаки единственного числа (SG) и именительного/винительного падежа (NOM/ACC) существительных и местоимений как немаркированные, и, соответственно, они не были размечены (Таблица 3 - строки 2, 3, 5, 6, 7, 9, 10 и др.). Отдельной строкой дается также литературный перевод данного высказывания. Такой формат делает корпус текстов доступным не только тем, кто владеет армянским языком, но и широкому кругу специалистов.

## 3. Сегментация устных нарративов и общие принципы транскрибирования

В вопросе сегментации устных нарративов и общих принципов транскрибирования явлений, которые относятся как к сегментному, так и к супraseгментному уровням, в целом, мы опирались на систему транскрибирования, разработанную и протестированную на русском корпусе «Рассказов о сновидениях» группой исследователей под руководством А.А. Кибрика и В.И. Подлеской, см.,

например, [4] (далее СТР - Система транскрибирования для русского языка). Основываясь на принципах, сформулированных в [4], мы приняли для армянского языка следующие решения.

При транскрибировании корпусного материала сохраняется аутентичность текста, в том числе, регистрируются заведомые отличия от стандартного языка (например, употребление формы 'a'=быть:PRS:3SG вместо ее литературного аналога 'ē', см. Таблица 3 – строки 4, 5, 6, 8), форм простых указательных местоимений ēs 'этот', ēd 'тот', ēn 'тот' вместо литературных ays, ayd, аун см. Таблица 3 – строка 3) и т.д.

Поскольку пунктуационная система армянского языка сильно отличается от стандартных систем пунктуации, используемых в европейских языках, в качестве исходной системы знаков препинания и их употребления взята система русского языка, практически в том виде, в котором она используется в СТР.

Используемый метод транскрибирования предполагает сегментацию текста на элементарные дискурсивные единицы (ЭДЕ), каждая из которых составляет отдельную строку. При сегментации текста на ЭДЕ базовыми являются два критерия:

- семантико-синтаксический - ЭДЕ представляет одну предикацию и
- просодический - ЭДЕ представляет собой одну интонационную единицу.

Однако, имеется ряд случаев, когда базовые два критерия сегментации не скоррелированы, и в результате, имеются ЭДЕ, которые больше или меньше одной клаузы или одной интонационной единицы (вынесенный топик, парцелляция, регуляторные фразы, фальстарт и т.д.; инфинитивные подлежащие, инфинитивные сентенциальные актанты, редупликация предиката, редукция одного из предикатов до статуса цитационного или другого маркера, быстрая подстановка и т.д.) Для каждого из подобных случаев разработаны конкретные последовательные правила, которые позволяют иметь систематизированный метод сегментации (более подробно см. [18]).

Так, нами были приняты специальные сегментационные решения, касающиеся представления в транскриптах нефинитных форм. Проведем краткий обзор нефинитных форм и принятые нами решения по их сегментации в армянском языке<sup>8</sup>.

График 1. Система независимых нефинитных форм в армянском языке



<sup>8</sup> В парадигме армянского глагола имеются 8 нефинитных форм, 4 из которых не имеют отдельного независимого употребления и выступают только в качестве компонентов в аналитических глагольных формах. Здесь рассматриваются независимые формы, поскольку именно они выступают в нефинитных конструкциях.

Нефинитные формы, функционирующие в качестве простого определения, а также формы инфинитива в номинативе рассматривались как одна предикация с управляющей единицей. Более неоднозначными оказались нефинитные конструкции, которые фигурировали в качестве деепричастия, предикативного компонента в придаточных предложениях и т.д. Для последних случаев в вопросе сегментации решающим фактором является наличие/отсутствие несущего акцента (подробнее см. [18]). При наличии акцента на нефинитной форме было принято решение, выделять нефинитную конструкцию в отдельную ЭДЕ.

В армянской глагольной системе, особенно в устном дискурсе, довольно распространено употребление сериальных конструкций. Как правило, в данных конструкциях статус двух глаголов разный, один глагол выступает в качестве основного элемента, который выражает базовую семантику и акцентируется, а второй глагол, который имеет более нейтральное значение, никогда не бывает акцентированным. В армянском языке данные конструкции не имеют видовременных ограничений, могут выступать как в синтетических, так и в аналитических глагольных формах. Однако, если при синтетических формах оба глагола морфологически оформлены одинаково, то при аналитических формах, второй вспомогательный глагол может опускаться. Таким образом, данные конструкции всегда представляют морфологическое и семантическое единство, и интонационно оформляются как одно фонетическое слово. В наших корпусах все сериальные конструкции рассматривались как одна предикация, и соответственно включались в одну ЭДЕ (см. Таблица 3 – строки 6, 8, 9).

Паузация, удлинённые гласные и согласные звуки. В данном формате транскрипции обязательной является подробная разметка всех типов пауз, а также регистрация всех случаев удлинённых согласных и гласных звуков. В качестве основы при разметке вышеупомянутых явлений были приняты принципы СТР – паузы обозначаются при помощи двух, трех или четырех точек в зависимости от длительности и транскрибируются следующим образом (подробнее, см. [4]):

|                         |                 |           |
|-------------------------|-----------------|-----------|
| менее 0.1 сек           | не обозначаются | —         |
| от 0.1 до 0.4 сек (вкл) | две точки       | ..(0.3)   |
| от 0.5 до 0.9 сек (вкл) | три точки       | ...(0.7)  |
| 1 сек и более           | четыре точки    | ....(1.2) |

Мы отмечаем несколько типов пауз, классифицируя их на основе трех критериев:

1. Позиция пауз, на основании этого критерия выделяются:
  - а) пограничные паузы (между ЭДЕ)
  - б) внутренние паузы (внутри ЭДЕ)
2. Качество пауз, на основании этого критерия выделяются:
  - а) заполненные паузы
  - б) незаполненные паузы.
3. Состав пауз, на основании этого критерия выделяются:
  - а) простые паузы, которые состоят только из одной заполненной или незаполненной паузы,
  - б) сложные паузы, которые состоят из более, чем одной заполненной или незаполненной паузы.

Имеются следующие варианты сложных пауз:

- заполненная пауза + незаполненная пауза,
- незаполненная пауза + заполненная пауза,
- заполненная пауза + незаполненная пауза + заполненная пауза,

- незаполненная пауза + заполненная пауза + незаполненная пауза.

Теоретически возможны также более сложные типы пауз, в которых будет многократное чередование заполненных и незаполненных пауз. Однако, как показывают наши корпуса, даже типы (с) и (d) встречаются редко, более сложные типы составных пауз нам не встретились.

Незаполненные паузы, включающие вдохи, выдохи, цоканье и т.д. отмечались надстрочным индексом <sup>h</sup> после скобок (в которых отмечается продолжительность пауз):

(1) КК-ARM R1-6<sup>9</sup>

**Пограничная, сложная пауза** (незаполненная пауза со вздохом + заполненная пауза + незаполненная пауза)

$..(1.3)^h \text{ } \overset{\circ}{\text{am}}(0.7) \dots(0.8) \text{ } \overset{\circ}{\text{c}}^h \text{karovac}^h \text{av} / \text{gnel},$   
 $\overset{\circ}{\text{c}}^h \text{karovac}^h \text{-av} \quad \text{gn-el}$   
 չ-կարողաց-ավ գն-ել  
 NEG-мочь-AOR:3 купить-INF  
 'не смог купить,'

**Заполненная пауза**

$k^h \text{ani vor ayn} == ..(0.2) \text{ } \overset{\circ}{\text{gin}} \text{ } \overset{\circ}{\text{sat}} \text{ } \overset{\circ}{\text{bar}} \text{ } \overset{\circ}{\text{zr}} \text{ } \overset{\circ}{\text{er}}.$   
 $k^h \text{ani} \quad \text{vor} \quad \text{ayn} \quad \overset{\circ}{\text{gin}} \text{-} \overset{\circ}{\text{a}} \quad \overset{\circ}{\text{sat}} \quad \overset{\circ}{\text{bar}} \text{ } \overset{\circ}{\text{zr}} \quad \overset{\circ}{\text{er}}$   
 քանի որ այն գին-ը շատ բարձր էր  
 сколько что это цена-DEF очень высокий быть:AUX:PST:3SG  
 потому что это эта цена была очень высокая.'

**Незаполненная, внутренняя пауза**

Используемый формат транскрипции предполагает также регистрацию всех заполненных пауз, в том числе обязательное указание их абсолютной продолжительности. В основном, способ заполнения пауз проявляет универсальные тенденции, однако, поскольку, в отличие от незаполненных пауз, заполненные паузы проявляют некоторые свойства сегментных единиц (ср. [19], где заполненные паузы считаются особыми словами), они имеют также ряд лингвоспецифических черт в каждом отдельном языке. Заполненные паузы иногда относят к классу междометий [21, 20, 19, 22, 23], что частично объясняет их типологически универсальные и лингвоспецифические свойства. При описании и создании инвентаря заполненных пауз, с одной стороны мы старались создать максимально единообразный способ представления, а с другой стороны, поскольку заполненные паузы, в отличие от лексических единиц, не имеют стандартизированного и канонизированного варианта и часто существенным здесь оказывается фонетический фактор, мы пытались выработать унифицированную систему представления. При обозначении разных заполненных пауз в русском, английском и итальянском языках – языках, которые, наряду с армянским, представлены в нашем четырехязычном корпусе – мы как правило, основывались на существующих традициях (русский язык – [24], английский язык – [25, 19, 26], итальянский – [27]). В арменистике супraseгментные явления описаны слабо, нет устойчивых традиций их отражения в транскрипции. В связи с этим для обозначения заполненных пауз (а также некоторых других супraseгментных явлений) в армянском языке мы составили свой инвентарь на основе материала

<sup>9</sup> Для каждого примера мы даем указание на его идентификационный номер в составе корпуса.



АКК и АКД. При выборе соответствующих обозначений, с одной стороны, мы старались сохранить основные относительно универсальные принципы обозначения заполненных пауз, которые имеются в традициях других языков и представляются типологически-релевантными [25, 19, 6, 24, 28], с другой стороны, сохранить особенности фонологической системы и графические традиции армянского языка. Как и в случае с остальными тремя языками-объектами, для армянского языка мы также использовали условную, а не фонетическую транскрипцию.

Таблица 2. Представление заполненных пауз в армянском языке

| Одноморовые        |                             | Полиморовые                 |        |
|--------------------|-----------------------------|-----------------------------|--------|
| гласно-заполненные | согласно-гласно заполненные | согласно-гласно заполненные |        |
| 1. əə              | 1. əv                       | 1.                          | əhət   |
| 2. əh              | 2. av                       | 2.                          | ətə    |
| 3. ah              | 3. ət                       | 3.                          | ətəh   |
|                    | 4. əht                      | 4.                          | ətəv   |
|                    | 5. aət                      | 5.                          | əvh    |
|                    | 6. həh                      | 6.                          | əvt    |
|                    | 7. hət                      | 7.                          | əvə    |
|                    |                             | 8.                          | əvət   |
|                    |                             | 9.                          | əvətəv |

Как показывает таблица, в составе заполненных пауз могут встречаться как гласные, так и согласные фонемы. Как будет показано далее, имеются ограничения, какие именно гласные и согласные могут выступать в качестве заполнителей. Основными являются гласные заполнители, которые могут выступать, как отдельно, так и с согласными. Кроме того, заполненные паузы могут иметь одноморовый характер (т.е. один гласный звук), и полиморовый (т.е. больше одного гласного, и всегда с согласным/и звуками) (ср. [19], где различаются неудлиненные и удлиненные заполненные паузы, которые характеризуются как одноморовые и двухморовые соответственно). Отметим, что прототипической заполненной паузой является одноморовая, полиморовые заполненные паузы являются более редким явлением.

## (2) KK-ARM T2-7

....(1.3)<sup>h</sup> əvətəv(1.2) ..(0.1) De-e(0.1) sra pat<sup>h</sup>osə č<sup>h</sup>ap<sup>h</sup>azanc<sup>h</sup> /pəz̄ ē,  
 de sra pat<sup>h</sup>os-ə č<sup>h</sup>ap<sup>h</sup>azanc<sup>h</sup> pəz̄ ē  
 դէ սրա պափոս-ը շափազանց պարզ է  
 ну это:GEN пафос-DEF чрезвычайно просто быть:СОР:PRS:3SG  
 'ну, пафос этого [рассказа] чрезвычайно прост'

Существенно, что набор гласных и согласных фонем-заполнителей совпадает во всех четырех языках, различие состоит в том, что качество данных фонем разное в зависимости от фонологической системы данного языка [6, 29, 30].

1. Гласные звуки
  - a. [a], [ə], [aə], [ə:]
2. Согласные звуки
  - a. сонорный [m]
  - b. гортанный [h]
  - c. фрикативный [v], гортанная смычка

Удлиненные звуки. В наших корпусах регистрируются все удлинения и их абсолютная продолжительность, однако, поскольку часто очень сложно разграничить продолжительность собственно звука в качестве компонента лексической единицы и продолжительность удлинения, то наши зарегистрированные величины продолжительности являются в какой-то мере относительными. В качестве удлиненного звука могут выступать как гласные, так и согласные. Однако, как показывают наши результаты, гласные звуки чаще удлиняются и, как правило, имеют большую продолжительность, чем согласные.

### (3) KK-ARM T1-1

...(0.6) /-yev-v(0.3) ..(0.2) əə(0.4) yeritasard /tʰamardə /-vorošum ē-ē(0.2) ...(0.5) əə(0.6) /gnel knoʒ hamar ..(0.1) /pʰokʰrik ..(0.2) /ba-aycʰ(0.2) ..(0.1) /harmaravet-t(0.2) əə(0.3) \avtomekʰena.

|     |            |             |             |                  |            |
|-----|------------|-------------|-------------|------------------|------------|
| yev | yeritasard | tʰamard-ə   | voroš-um    | ē                | gn-el      |
| li  | երիտասարդ  | տղամարդ-ը   | որոշում     | է                | գն-ել      |
| и   | молодой    | человек-DEF | решать-IPFV | быть:AUX:PRS:3SG | купить-INF |

|              |       |           |       |            |             |
|--------------|-------|-----------|-------|------------|-------------|
| kn-oʒ        | hamar | pʰokʰrik  | baycʰ | harmaravet | avtomekʰena |
| կն-ոջ        | համար | փոքրիկ    | բայց  | հարմարավետ | ավտոմեքենա  |
| жена-GEN/DAT | для   | маленький | но    | удобный    | машина      |

'и молодой человек решил купить для жены маленькую, но удобную машину.'

Тоны/акценты. Данная система транскрибирования предполагает также подробную разметку движения тона и акцентоносителей, поскольку они являются основными просодическими маркерами и выражают семантику дискурса. Место акцента и тона определяется на перцептивном уровне и с помощью программы Speech Analyzer<sup>10</sup>. Был использован следующий набор тонов: / акцент с восходящим тоном, \ акцент с нисходящим тоном, /- акцент с восходяще-ровным тоном, \- акцент с нисходяще-ровным тоном, - акцент с ровным тоном, / акцент с восходяще-нисходящим тоном, \ акцент с нисходяще-восходящим тоном.

Расстановка тонов в наших корпусах преследовала сугубо вспомогательные цели, а именно, была нацелена на максимальную регистрацию информации, которая эксплицировала бы интерпретацию хезитационных явлений. Поэтому, несмотря на то, что мы старались размечать все основные тоновые контуры, эта часть разметки не была сделана с достаточной полнотой и детализацией, тем более что для армянского языка отсутствует комплексное описание интонационной системы, что делает нашу задачу особенно сложной.

Пунктуационные знаки/ Орфография. Употребление пунктуационных знаков в СТР иконическое, выработаны специальные правила, которые регулируют их распределение как средств обозначения типов иллокуции и сигналов завершенности/незавершенности иллокуции. В частности для обозначения границы иллокуций могут быть использованы следующие знаки: ‘.’ – конец «сообщения», ‘.’ – многоточие, ‘!’ – директив, ‘?’ - интеррогатив, ‘~’ – оборванное высказывание, ‘@’ - вокатив.

В наших корпусах иконически отображается смех (☺), и кашель (☹) (см. Таблица 3 – строка 14], остальные паралингвистические явления (например, смех, кашель, цоканье и т.д.) отмечаются в форме комментариев.

<sup>10</sup> Speech Analyzer 1.5 Speech Manager 1.5, IPA Help 2.0, and AudioCon 3.2 Copyright (c) 2000 SIL International.

При распознавании текста те слова, для которых транскрайбер допускает несколько вариантов интерпретации, даются во всех вариантах через вертикальную черту в угловых скобках. Как и в СТР, с помощью подчеркиванием пунктиром отмечается нестандартная редукция слов.

В качестве иллюстрации вышеизложенного, приведем образец затранскрибированного армянского устного нарратива.

Таблица 3. Образец затранскрибированного текста

**KK-ARM R1-5**

| <i>t</i>    | <i>N</i>  | <i>TEXT</i>   |
|-------------|-----------|---|
| <b>1.7</b>  | <b>1.</b> | <b>..(0.4)<sup>n</sup> Tak /uremn,</b>  |
|             |           | tak uremn<br>տակ ուրեմն<br>так значит<br>Так, значит,   |
| <b>2.7</b>  | <b>2.</b> | <b>əə(0.5) ...(0.5)<sup>n</sup> mi garnanayi-in(0.2) ..(0.1) paɣsaʔ arevot /ōr ēr,</b>  |
|             |           | mi garnanayin paɣsaʔ arevot ōr ēr<br>մի գարնանային պայծառ արևոտ օր էր<br>один весенний светлый солнечный день быть:AUX:PST:3SG<br>был весенний светлый солнечный день,  |
| <b>5.9</b>  | <b>3.</b> | <b>u ēd /amusinneri-i(0.3) ..(0.3)<sup>n</sup> əm(0.3) ...(0.6)<sup>n</sup> amusnakan /taredarɜn ēr,</b>  |
|             |           | u ēd amusin-ner-i amusnakan taredarɜ-n ēr<br>ու էդ ամուսին-ներ-ի ամուսնական տարեդարձ-ն էր<br>и это супруг-PL-GEN/DAT супружеский юбилей-DEF быть:AUX:PST:3SG<br>и, это самое, был юбилей свадьбы у этих супругов,   |
| <b>9.6</b>  | <b>4.</b> | <b>u inč<sup>n</sup> pes-s(0.5) ..(0.4)<sup>n</sup> petk<sup>n</sup> a /liner,</b>  |
|             |           | u inč <sup>n</sup> pes petk <sup>n</sup> a lin-er<br>ու ինչպես պետք ա լին-եր<br>и как должен быть:AUX:PRS:3SG быть-SBJV:PRS:3SG<br>и как полагается,  |
| <b>11.5</b> | <b>5.</b> | <b>/amusinə ...(0.5)<sup>n</sup> /gnum a ..(0.1) /nver ..(0.1) p<sup>n</sup>ntrelu ir knoʒ \hamar.</b>  |
|             |           | amusin-ə gn-um a nver p <sup>n</sup> ntrelu-er<br>ամուսին-ը գն-ում ա նվեր փնտրել-ու<br>супруг-DEF идти-IPFV быть:AUX:PRS:3SG подарок искать-INF-GEN/DAT<br>ir knoʒ hamar<br>իր կնոջ համար<br>он:GEN жена-GEN/DAT для<br>супруг пошел искать подарок для своей жены. |
| <b>14.7</b> | <b>6.</b> | <b>..(0.4)<sup>n</sup> Gnum a tarber ɣanut<sup>n</sup> ner a man /galis,</b>  |
|             |           | gn-um a tarber ɣanut <sup>n</sup> -ner a<br>գն-ում ա տարբեր խանութ-ներ ա<br>идти-IPFV быть:AUX:PRS:3SG разный магазин-PL быть:AUX:PRS:3SG<br>man gal-is<br>ման գալ-իս<br>искать:NVP прийти-IPFV: VVP<br>Ходит ищет в разных магазинах,                              |
| <b>16.9</b> | <b>7.</b> | <b>..(0.1)<sup>n</sup> voč<sup>n</sup> mi ban č<sup>n</sup>i /gtnum,</b>  |
|             |           | voč <sup>n</sup> mi ban č <sup>n</sup> -i gtn-um<br>ոչ մի բան չ-ի գտն-ում<br>не один:INDF вещь NEG-быть:AUX:PRS:3SG найти-IPFV<br>ничего не находит,  |
| <b>18.2</b> | <b>8.</b> | <b>....(1.2)<sup>n</sup> heto galis a ērexek<sup>n</sup>i het /ɣorhrdak<sup>n</sup>um a,</b>  |
|             |           | heto gal-is a ērex-ek <sup>n</sup> -i het<br>հետո գալ-իս ա էրեխ-եր-ի հետ<br>потом прийти-IPFV быть:AUX:PRS:3SG дитя-PL-GEN/DAT с  |

|             |            |   |                          |                  |   |
|-------------|------------|---|--------------------------|------------------|---|
|             |            | χorhrdak <sup>h</sup> -um a   |                          |                  |   |
|             |            | խորհրդակցում ա  |                          |                  |   |
|             |            | советоваться-IPFV быть:AUX:PRS:3SG  |                          |                  |   |
|             |            | потом приходит посоветоваться с детьми,   |                          |                  |   |
| <b>21.4</b> | <b>9.</b>  | <b>..(0.7)<sup>h</sup> \-yereχanerə-ə(0.3) ..(0.2) χorhurd en talis gnal mek<sup>h</sup>ena /afnel,</b>                             |                          |                  |   |
|             |            | yereχa-ner-ə  | χorhurd                  | en               | tal-is gn-al mek <sup>h</sup> ena afn-el          |
|             |            | երեխա-ներ-ը   | խորհուրդ                 | են               | տալ-իս գն-ալ մեքենա աֆն-ել                        |
|             |            | дитя-PL-DEF   | совет                    | быть:AUX:PRS:3PL | дать-IPFV иди-INF машина купить-INF               |
|             |            | дети советуют пойти купить машину,  |                          |                  |   |
| <b>25.0</b> | <b>10.</b> | <b>k<sup>h</sup>ani vor ..(0.4)<sup>h</sup> ayd /mek<sup>h</sup>enan ==</b>   |                          |                  |   |
|             |            | k <sup>h</sup> ani  | vor                      | ayd              | mek <sup>h</sup> ena-n                            |
|             |            | քանի որ այդ   | մեքենա-ն                 |                  |   |
|             |            | сколько что это машина-DEF  |                          |                  |   |
|             |            | поскольку эта машина ==   |                          |                  |   |
| <b>26.5</b> | <b>11.</b> | <b>yev-v(0.3) ..(0.1) /patčafabanum en,</b>   |                          |                  |   |
|             |            | yev   | patčafaban-um            | en               |   |
|             |            | և   | պատճառաբանում            | են               |   |
|             |            | и аргументировать-IPFV быть:AUX:PRS:3PL   |                          |                  |   |
|             |            | и аргументируют,  |                          |                  |   |
| <b>27.5</b> | <b>12.</b> | <b>vor yev anhražešt ēr yev irenc<sup>h</sup> tan /hamar,</b>   |                          |                  |   |
|             |            | vor   | yev                      | anhražešt        | ēr yev irenc <sup>h</sup> tan hamar               |
|             |            | որ և  | անհրաժեշտ էր             | և                | իրենց տան համար                                   |
|             |            | что и необходимо быть:AUX:PST:3SG и они:GEN дом:GEN для   |                          |                  |   |
|             |            | [это то] что и необходимо и для их дома,  |                          |                  |   |
| <b>29.4</b> | <b>13.</b> | <b>yev ..(0.1) mayrikə šat /kuraχana,</b>   |                          |                  |   |
|             |            | yev   | mayrik-ə                 | šat              | k-uraχan-a  |
|             |            | և   | մայրիկ-ը                 | շատ              | կ-ուրախան-ա                                       |
|             |            | и мать-DEF очень COND-радоваться-PRS:3SG  |                          |                  |   |
|             |            | и мама очень обрадуется,  |                          |                  |   |
| <b>31.0</b> | <b>14.</b> | <b>...(0.5)<sup>h</sup> laypisov ...(0.7) ⊕ əntanik<sup>h</sup>i hayrə gnum ē \mek<sup>h</sup>enan= ..(0.1) /χanut<sup>h</sup>,</b> |                          |                  |   |
|             |            | laypisov  | əntanik <sup>h</sup> -i  | hayr-ə           | gn-um ē mek <sup>h</sup> ena-n χanut <sup>h</sup> |
|             |            | այսպիսով  | ընտանիք-ի                | հայր-ը           | գնում է մեքենա-ն խանութ                           |
|             |            | так   | семья                    | отец-DEF         | идти-IPFV быть:AUX:PRS:3SG машина-DEF магазин     |
|             |            | таким образом, отец семьи идет в машину= в магазин,   |                          |                  |   |
| <b>34.4</b> | <b>15.</b> | <b>vorteχ mek<sup>h</sup>enaner en /vačafum,</b>  |                          |                  |   |
|             |            | vorteχ  | mek <sup>h</sup> ena-ner | en               | vačaf-um  |
|             |            | որտեղ   | մեքենա-ներ               | են               | վաճառում  |
|             |            | где   | машина-PL                | быть:AUX:PRS:3PL | продавать-IPFV                                    |
|             |            | где продают машины,   |                          |                  |   |
| <b>36.0</b> | <b>16.</b> | <b>...(0.6)<sup>h</sup> aəm(0.6) canot<sup>h</sup> anum ē /gnerin,</b>  |                          |                  |   |
|             |            | canot <sup>h</sup> an-um  | ē                        | gn-er-i-n,       |   |
|             |            | ծանոթանում  | է                        | գն-եր-ի-ն        |   |
|             |            | ознакомиться-IPFV быть:AUX:PRS:3SG цена-PL-GEN/DAT-DEF  |                          |                  |   |
|             |            | узнает ценами,  |                          |                  |   |
| <b>38.4</b> | <b>17.</b> | <b>yev k<sup>h</sup>ani vor č<sup>h</sup>i karovanum ..(0.4)<sup>h</sup> gnel /ayn,</b>   |                          |                  |   |
|             |            | yev   | k <sup>h</sup> ani       | vor              | č <sup>h</sup> -i karovan-um gn-el ayn            |
|             |            | և   | քանի որ չ-ի              | կարոճանում       | գն-ել այն   |
|             |            | и сколько что NEG-быть:AUX:PRS:3SG мочь-IPFV купить-INF это   |                          |                  |   |
|             |            | и поскольку не может ее купить,   |                          |                  |   |
| <b>40.5</b> | <b>18.</b> | <b>vorošum ē gnel χavalik<sup>h</sup> /mek<sup>h</sup>ena,</b>  |                          |                  |   |
|             |            | voroš-um  | ē                        | gn-el            | χavalik <sup>h</sup> mek <sup>h</sup> ena         |
|             |            | որոշում   | է                        | գն-ել            | խաղալիք մեքենա                                    |
|             |            | решить-IPFV быть:AUX:PRS:3SG купить-INF игрушка машина  |                          |                  |   |
|             |            | решает купить игрушечную машину,  |                          |                  |   |
| <b>42.2</b> | <b>19.</b> | <b>...(0.9)<sup>h</sup> əh(0.2) χavalik<sup>h</sup> mek<sup>h</sup>enan tun ē /berum,</b>   |                          |                  |   |
|             |            | χavalik <sup>h</sup>  | mek <sup>h</sup> enan    | tun              | ē ber-um  |

|             |            |  |
|-------------|------------|--|
|             |            | խաղալիք մեքենան տուն է բերում<br>игрушка машина дом быть:AUX:PRS:3SG принести-IPFV<br>несет игрушечную машину домой,   |
| <b>44.6</b> | <b>20.</b> | <b>yev-v(0.2) yerb vor ..(0.2)<sup>h</sup> nvernerə p<sup>h</sup>oḡanakelu /žamanak ..(0.1) talis ē ir /mek<sup>h</sup>enaṅ,</b>   |
|             |            | yev yerb vor nver-ner-ə p <sup>h</sup> oḡanak-el-u žamanak<br>և երբ որ նվեր-ներ-ը փոխանակ-ել-ու ժամանակ<br>и когда что подарок-PL-DEF обменять-INF-GEN/DAT время<br>tal-is ē ir mek <sup>h</sup> ena-n<br>տալ-իս է իր մեքենա-ն<br>дать-IPFV быть:AUX:PRS:3SG он:GEN машина-DEF<br>и когда во время обмена подарков отдает свою машину, |
| <b>48.2</b> | <b>21.</b> | <b>kinə šat ==</b>   |
|             |            | kin-ə šat<br>կին-ը շատ<br>жена-DEF очень<br>жена очень,  |
| <b>48.6</b> | <b>22.</b> | <b>...(0.7)<sup>h</sup> əm(0.4) ..(0.3)<sup>h</sup> aydk<sup>h</sup>an durə č<sup>h</sup>i galis /knožə,</b>   |
|             |            | aydk <sup>h</sup> an el dur-ə č <sup>h</sup> -i gal-is kn-ož-ə<br>այդքան էլ դուր-ը չ-ի գալ-իս կն-ոջ-ə<br>столько тоже нравиться-DEF:NVP NEG-быть:AUX:PRS:3SG идти-IPFV:VVP жена-GEN/DAT-DEF<br>жене не очень нравится,   |
| <b>51.5</b> | <b>23.</b> | <b>sakayn lyereḡanerə šat en luraḡanum</b>   |
|             |            | sakayn yereḡa-ner-ə šat en uraḡan-um<br>սակայն երեխա-ներ-ը շատ են ուրախանում<br>но дитя-PL-DEF очень быть:AUX:PRS:3PL радоваться-IPFV<br>но дети очень радуются.   |

## Литература

1. Khurshudyan V. *Hesitation Markers in Oral Narratives in Armenian and Italian*. // VII International Conference on Cognitive Modeling in Linguistics, Varna, Bulgaria, 2004, pp. 125-135.
2. Хуришудян В. Г. Показатели хезитации в устных нарративах армянского языка в сопоставлении с итальянским. // Четвертая типологическая школа. Международная школа по лингвистической типологии и антропологии. Цахкадзор, Армения, 2005 г.
3. Хуришудян В. Г. Экспериментальное исследование хезитации в разно-структурных языках. // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции "Диалог'2005". М.: Наука, 2005.
4. Кибрик А.А., Подлеская В.И. К созданию корпусов устной русской речи: принципы транскрибирования. // Научно-техническая информация. Серия 2: Информационные процессы и системы, 2003 № 10, сс. 5-11.
5. DuBois J. W., Schuetze-Coburn S., Cumming S., Paolino D. *Outline of discourse transcription // Talking data: Transcription and coding in discourse research / Hillsdale, N. J., 1993, pp. 45-89.*
6. Edwards J. A. *The transcription of discourse // The handbook of discourse analysis. Oxford: Blackwell, 2001, pp. 321-348.*
7. MacWhinney B. *The CHILDES Project: tools for analyzing talk, v. I-II (third edition), Lawrence Erlbaum Associates, New Jersey, 2000.*
8. MacWhinney B. & Snow C. *The Child Language Data Exchange System. // Journal of Child Language, # 12, 1985, pp. 271-296.*
9. Кибрик А.А. Анализ дискурса в когнитивной перспективе. Диссертация на соискание ученой степени доктора филологических наук. М. 2003.
10. Гармаиш Н.Г. Влияние хезитации на организацию устного детского дискурса. Диссертация на соискание ученой степени кандидата филологических наук. М. 1999.

11. Подлеская В. И., Хуришудян В. Г. О лексических маркерах хезитации в спонтанной речи: уроки армянского. // *Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции "Диалог'2006"*, М. 2006. сс. 429-439.
12. Маркосян А.С. Психолингвистические особенности разговорной речи. Диссертация на соискание ученой степени кандидата филологических наук. М. 1983.
13. SAPP (1981). *Ճամբակակի՝ հայերենի զօսակ՝ական լեզուն* [Современная армянская разговорная речь]. Ереван.
14. Sanjian A. K. *The Armenian alphabet. // The world's writing systems. New York & Oxford: Oxford University Press. 1996, pp. 356-363.*
15. Гюбиман Г., *Հայագիտական usumnasirut՝yunner* [Арменоведческие исследования]. Ереван. 2004, pp. 56-75.
16. Кодзасов С.В., Кривнова О.Ф. *Общая фонетика. М., 2001.*
17. *The Leipzig Glossing Rules, version of September 2004, <http://www.eva.mpg.de/lingua/pdf/LGR04.09.21.pdf>.*
18. Кибрик А.А. Подлеская В.И. Проблема сегментации устного дискурса и когнитивная система говорящего // В.Д.Соловьев (ред.) *Когнитивные исследования. Вып. 1. Москва: Институт психологии РАН, 2006, С.138-158.*
19. Clark H. H. & Jean E. Fox Tree. *Using uh and um in Spontaneous Dialog. // Cognition, 2002, #84, pp. 73-111.*
20. Godfrey J.J., Holliman E.C., McDaniel J., *Switchboard: Telephone speech corpus for research and development. // Conference on Acoustics, Speech and Signal Processing. San Francisco, 1992, pp. 517-520.*
21. James D. *Some aspects of the syntax and semantics of interjections. // Eighth Regional Meeting of the Chicago Linguistic Society, Chicago: Linguistics Department, University of Chicago, 1972, pp. 162-172.*
22. Clark H. H. *Pragmatics of language performance. // Handbook of pragmatics. Oxford: Blackwell, 2004, pp. 365-382.*
23. Engel D., Charniak E., Johnson M. *Parsing and Disfluency Placement, // Conference on Empirical Methods in Natural Language Processing, Philadelphia, July 2002, pp. 49-54.*
24. Земская Е. А., Китайгородская М. В., Ширяев Н. Е. *Русская разговорная речь. Общие вопросы. Словообразование. Синтаксис. М.: Наука, 1981.*
25. DuBois, J. W., Cumming S., Scheutze-Coburn S., Paolino D. *Discourse Transcription. UCSB, 1992.*
26. Ward N. *Issues in the Transcription of English Conversational Grunts. // First (ACL) SIGdial Workshop on Discourse and Dialog, 2000, pp. 29-35.*
27. Cresti E. *Corpus di italiano parlato, v. I-II, Accademia della Crusca, Firenze, 2000.*
28. Shriberg E.E. *Preliminaries to a Theory of Speech Disfluencies. Doctoral dissertation. University of California at Berkeley. 1994.*
29. Kawamori M., Kawabata T., Shimazu A. *Discourse markers in spontaneous dialogue: a corpus based study of Japanese and English. // ACL98 Workshop of Discourse Relations and Discourse Markers, 1998, pp. 93-99.*
30. Swerts M., Wichmann A., Beun R.-J. *Filled Pauses as Markers of Discourse Structure. // ICSLP, Philadelphia, USA, 1996.*