



Текст курсовой

# Инвентарь форм рук классификаторных предикатов в РЖЯ

Чевелева Анастасия, БКЛ192, ФикЛ, НИУ ВШЭ

Научный руководитель: Толдова Светлана Юрьевна

Консультант: Клезович Анна Геннадьевна

30/05/2022

# Русский Жестовый Язык (РЖЯ)

- Полноценная языковая система
- Носители: глухие, слабослышащие, члены их семей
- Особенности: визуальная модальность, иконичность (Pierce 1931)
- **Классификаторы** (classifiers, depicting signs) – форма руки ассоциирована с некоторым классом объектов (Zwitserlood 2012)



Рис. 1. clf:человек:двигаться  
'человек идет' (Корпус РЖЯ)



Рис. 2. ножницы.резать clf:плоский.предмет:держат  
'резать ножницами листок бумаги' (Корпус РЖЯ)

# Постановка проблемы, цели и задач

- Классификаторы крайне важны для обработки ЖЯ, т.к. **частотны**. Однако они **частично лексикализованные**, что создает трудности для языковых моделей;
- Идея парсинга ЖЯ: seq2seq перевод звучащего – глоссы жестового;
- Для упрощения парсинга и генерации классификаторов предлагается использовать **ограниченный инвентарь конфигураций** (форм рук);
- На данный момент не существует исследования РЖЯ, покрывающего все классификаторы;
- Существующая в корпусах разметка не унифицирована и не позволяет быстро автоматически провести анализ конфигураций.

**Цель:** определить общий инвентарь рук классификаторных предикатов РЖЯ

## **Задачи:**

1. Проанализировать существующие разметки классификаторных предикатов;
2. Разработать систему разметки классификаторных предикатов;
3. Переразметить классификаторные предикаты в имеющемся корпусе данных;
4. Проанализировать конфигурации рук в полученных конструкциях.

# Проблема определения классификаторов

Основные виды классификаторов (Supalla 1982, 1986):

- **Size-and-shape-specifiers (SASSes)** – внешние признаки
  - Static – объект целиком
  - Tracing – движение, описывающее контур объекта
- **Semantic classifiers** – семантические классы (одушевленные двуногие объекты, длинные тонкие предметы) (но опираются на внешние признаки!)
- **Instrumental classifiers**
  - Сам объект ~ static SASSes/semantic classifier
  - Рука в момент манипуляции с объектом – **handling classifiers**
- **Body(-part) classifiers** (но body classifiers → referential/role shift)

# Проблема определения классификаторных предикатов

**Классификаторные предикаты** – классификатор + глагол, в предложении играют роль предиката

Подходы к анализу:

- Классификатор – аффикс, глагол – корень (Supalla 1986 для SASSes)
  - Классификатор ~ согласование (Benedicto & Bretari 2004)
- Два корня
  - Равноправные (Zwitserlood 2003 для части body-part classifiers)
  - Классификатор – инкорпорированный (Meir 2001)
- Единый жест (Liddell 2003) + проблема ‘frozen’ sings – лексикализованные жесты

# Исследования классификаторов в РЖЯ

- (Kyuseva 2020): SASSes\* – движение не может определять перемещение или локализацию в пространстве (16 конфигураций)
- (Христофорова 2017): semantic classifiers (14 конфигураций)

**Объект данного исследования:** классификаторные предикаты

- ~~SASSes\*~~, т.к. аргумент/атрибутив
- ~~'Frozen' signs~~, т.к. не представляют проблемы для языковой модели, можно записать в словарь
- ~~Body classifiers~~, т.к. нет слоя разметки для тела
- Static SASSes (по (Supalla 1982)) + semantic + handling + body-part
  - Глаголы локализации, движения, перемещения
  - Другие глаголы
- Рассматриваем как одноручные, без немануального (не\_руки) компонента

# Данные

Корпус компании «Адаптис»:

- Видео (NDA) + eaf-разметка
  - ПР-гloss – glossы правой руки
  - ЛР-гloss – glossы левой руки
  - Перевод – перевод на русский (звучащий) язык
- 6089 предложений (нарративы, сказки, элицитация)

Для анализа: 5 видео из топа по уникальным вхождениям классификаторов и общему числу вхождений. Всего к переразметке: **888 вхождений**.

## Изначальная разметка

## Проблемы существующей разметки:

- Непоследовательно *clf:объект:движение* (только сущ, только глагол, только наречие)
- Что такое классификатор?
- Что такое объект? Что такое движение?
- Фон и Фигура: одинаково

	00:01:23.000	00:01:24.000	00:01:25.000	00:01:26.000	00:01:27.000	00:01:28.000	00:01:29.000
PR-гloss [710]	Недвое	часто	приходить	clf:берег	плавание	clf:c.берега.прыгать.в.воду%	
LP-гloss [443]			приходить	clf:берег	плавание	clf:c.берега.прыгать.в.воду	
Перевод [89]	Они часто приходят на берег и плавают, прыгая в воду.						



Существующие разметки:

- Компромисс: **clf:референт:Nº формы:глагол**

- |                      | 00:01:23.000  | 00:01:24.000 | 00:01:25.000 | 00:01:26.000 | 00:01:27.000 | 00:01:28.000                 | 00:01:29.000 |
|----------------------|---|--------------|--------------|--------------|--------------|------------------------------|--------------|
| PR-гloss<br>[710]    | №двое   | часто        | приходить    | clf:берег    | плавание     | clf:c.берега.прыгать.в.воду% |              |
| PR-гloss_clf<br>[48] |   |              |              |              |              | clf:медведь:51:прыгать.в     |              |
| LP-гloss<br>[444]    |   |              | приходить    | clf:берег    | плавание     | clf:c.берега.прыгать.в.воду  |              |
| LP-гloss_clf<br>[46] |   |              |              |              |              | clf:берег:71:быть            |              |
| Перевод<br>[89]      | Они часто приходят на берег и плавают, прыгая в воду. |              |              |              |              |                              |              |

# Решение сложных кейсов с помощью новой разметки

- **Нет(?) глагола:** «(под снегом лед,) а под ним – вода» ~ clf:вода:87:течь, поскольку движение не про локацию
- **Semantic classifier V.S. Body-part:** «медвежонок» (sem), «медвежонок» (плывет на бочке; br), «медвежонок» (ступил на песок и обжегся; br), «лапы» (опустил в воду; br) ~ *медвежонок, медвежонок(?), лапа(?), лапа* – то, что в контексте-переводе
- **Взаимное расположение рук:** идти на четырех лапах V.S. два медведя – честно не смотрим
- **Относительность движения:** «медвежонок плывет к бочке» и «медвежонок плывет на бочке» – разные глаголы *плыть* и *плыть.на*

# Полученный инвентарь форм (топ по частоте)

162 различных вхождения **clf:референт:№формы:глагол**, 27 форм

Формы, закод. по (Клезович 2019)	Формы итоговые (в данной работе)	Формы по (Христофорова 2017)	Комментарий различий с (Христофорова 2017)	Примеры класс-х констр-й	Число конструкций
71	B	B		льдина:быть	30
59	2b	2b		медвежонок:лежать	22
1	A	A	+handling	человек:упасть	15
142	E	E	+ семантика маленького кусочка	виноград:проглотить	12
87	5	5		медвежонок:обжечься	12
95	500	500		желудок:быть	9

# Полученный инвентарь форм: сравнение с (Христофорова 2017)

- Формы В~Ф~Фb
  - 'Большой плоский объект' (Христофорова 2017): поезд:ехать, дельфин:плыть
  - **Body-part:** рука:хлопать, лапа:опускать
- Формы 2, 2b, 1, 1b
  - 'одушевленные' (там же): медвежонок, человек, орел
- Формы 5, 500
  - 'Объемные объекты неопределенной формы' (там же): еда, бегемот
  - **Static SASS:** здание, холм
  - **Body-part:** лапа, зубы

# Полученный инвентарь форм: сравнение с (Христофорова 2017)

- А
  - (там же): голова, человек
  - **Handling:** дверца.духовки:открывать, куртка:надевать
- С
  - (там же): бочка
  - **Handling:** бутылка:держатъ, сэндвич:есть
- h
  - **Handling:** таз:обливаться, ботинок:надеть

**Handling, которых не было, но точно существуют:** 500, E

Итого (с учетом алломорфов): **13 форм**

# Результаты

- Выделено 13 форм
  - Недостаточно для покрытия всех (меньше, чем у (Христофорова 2017))
  - Но выделены те, которых не было раньше + другие подвиды
- Улучшение:
  - Больше данных – сложно собирать
  - Лучше разметка – либо усложнение, либо еще менее информативно
  - ! инструменты по обработке видео !
- Польза созданной разметки для seq2seq моделей: векторизация референта и/или глагола
  - + модели кластеризации (сравнить классы с полученными в исследованиях)
  - + модели предсказания (например, если референт хорошо помогает предсказать форму, то требование к большой модели на снятие референции)

# Список литературы

- Дрозденко 2018 – Дрозденко О. С. Принципы разработки базы жестов РЖЯ со специфическим ротовым компонентом. МГУ, М., 2018.
- Клезович 2019 – Клезович А. Г. Автоматическое извлечение фонемного инвентаря русского жестового языка. ВШЭ, М., 2019.
- Корпус РЖЯ – Корпус русского жестового языка [Электронный ресурс]. Руководитель проекта С. И. Буркова, Новосибирск, 2012–2015.  
Режим доступа: <http://rsl.nstu.ru/> (дата обращения: 23.05.2022)
- Кюсева 2018 – Кюсева М. В. Физические свойства в русском жестовом языке в типологическом освещении. ВШЭ, М., 2018.
- Христофорова 2017 – Христофорова Е. А. Инвентарь семантических классификаторов русского жестового языка. РГГУ, М., 2017.
- Baker & Padden 1978 – Baker, C., Padden, C. A. (1978). Focusing on the Non-manual Components of ASL. In Siple, Patricia (ed.), *Understanding Language through Sign Language Research*. New York: Academic Press, pp. 27–57.
- Benedicto & Brentari 2004 – Benedicto, E., Brentari D. (2004). Where Did All the Arguments Go? Argument-changing Properties of Classifiers in ASL. In *Natural Language and Linguistic Theory* 22, pp. 743–810.
- Boyes-Braem & Sutton-Spence 2001 – Boyes-Braem, P., Sutton-Spence, R. (Eds.). (2001). *The hands are the head of the mouth: The mouth as articulator in sign languages*. Signum Verlag.
- Crasborn et al. 2020 – Crasborn, O., Zwitserlood, I., van der Kooij, E., Bank R. (2020, July 4). *Annotation Conventions for the Corpus NGT, version 4*. Radboud University, Nijmegen.
- Dixon 2004 – Dixon, R.M.W. (2004). Adjective classes in typological perspective. In Dixon R.M.W., Aikhenvald A.Y. (Eds.) *Adjective classes. A cross-linguistic typology*. Oxford University Press. pp. 1–49.
- ELAN – Hellwig, B., Hulsbosch, M., Somasundaram, A., Tacchetti, M., Geerts, J. (2018). *ELAN — Linguistic Annotator: version 6.3. Manual updated on 2022-01-24*. Retrieved May 26, 2022, from <https://www.mpi.nl/corpus/manuals/manual-elan.pdf>
- Filhol & McDonald 2020 – Filhol, M., McDonald, J. (2020, May). The synthesis of complex shape deployments in sign language. In *Proceedings of the 9th workshop on the Representation and Processing of Sign Languages*, pp. 61–68.
- Hanke 2004 – Hanke, T. (2004). HamNoSys — Representing sign language data in language resources and language processing contexts. In O. Streiter & C. Vettori (Eds.), *Workshop proceedings: Representation and processing of sign languages* (pp. 1–6). ELRA.

# Список литературы

- Hanke 2010 – Hanke, T. (2010). HamNoSys 4 Handshapes Chart. Drawings by H. Zienert, O. Jeziorski, A. Hanß. Retrieved from [https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt\\_pdf/HamNoSys\\_Handshapes.pdf](https://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/HamNoSys_Handshapes.pdf)
- Johnston 2019 – Johnston, T. (2019, August). Auslan Corpus Annotation Guidelines. Auslan Corpus.
- Jurafsky & Matrin 2022 – Jurafsky, D., Martin, J. H. (2022, January 12). Speech and Language Processing. 3rd Edition draft. Retrieved May 22, 2022, from [https://web.stanford.edu/~jurafsky/slp3/ed3book\\_jan122022.pdf](https://web.stanford.edu/~jurafsky/slp3/ed3book_jan122022.pdf)
- Kimmelman 2018 – Kimmelman, V. (2018). Basic argument structure in Russian Sign Language. In *Glossa: a journal of general linguistics* 3(1): 116, pp. 1–39.
- Kyuseva 2020 – Kyuseva, M. (2020). Size and shape specifiers in Russian Sign Language: a morphological analysis (PhD thesis). Retrieved from <http://hdl.handle.net/11343/241476>
- Liddell 2003 – Liddell, S. K. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge University Press.
- Mandel 1977 – Mandel, M. (1977). Iconic Devices in American Sign Language. In Friedman, Lynn A. (ed.), *On the Other Hand: New Perspectives on American Sign Language*. New York: Academic Press, pp. 57–107.
- McBurney 2012 – McBurney, S. (2012). History of sign languages and sign language linguistics. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign Language: An International Handbook* (pp. 909–949). Mouton De Gruyter.
- Meir 2001 – Meir, I. (2001). Verb Classifiers as Noun Incorporation in Israeli Sign Language. In Booij, Gerard/Marle, Jacob van (eds.), *Yearbook of Morphology 1999*. Dordrecht: Kluwer, 299–319.
- Pierce 1931 – Pierce, C. S. (1931). *Collected papers of Charles Sanders Pierce*. (C. Hartshorne & P. Weiss, Eds.). Cambridge, MA: Harvard University Press.
- Supalla 1982 – Supalla, T.R. (1982). *Structure and Acquisition of Verbs of Motion and Location in American Sign Language*. San Diego: PhD Thesis, UCSD.
- Supalla 1986 – Supalla, T. R. (1986). The Classifier System in American Sign Language. In C. Craig (Ed.), *Noun Classes and Categorization* (pp. 181–214). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Talmy 1985 – Talmy, L. (1985). Lexicalization Patterns: Semantic Structure in Lexical Forms. In Shopen, T.E. (Ed.) *Language Typology and Syntactic Description*, pp. 57–149.
- Zaki et al. 2014 – Zaki, M. J., Meira Jr, W., & Meira, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.
- Zwitserlood 2003 – Zwitserlood, I. (2003). *Classifying hand configurations in Nederlandse Gebarentaal (Sign Language of the Netherlands)*. Utrecht: LOT.
- Zwitserlood 2012 – Zwitserlood, I. (2012). Classifiers. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign Language: An International Handbook* (pp. 158–186). Mouton De Gruyter.



# Спасибо за внимание!

**CREDITS:** This presentation template was created  
by **Slidesgo**, including icons by **Flaticon**, and  
infographics & images by **Freepik**

Please keep this slide for attribution