

Referential devices of episode distinction on the material of the Russian Clinical Pear Stories Corpus

E. Ivtushok, K. Shulgina, S. Toldova, N. Akutina
Corresponding author: e.ivtushok@gmail.com



MATERIALS

RUSSIAN CLIPS CORPUS

Russian Clinical Pear Stories:

- Consists of Pear Stories retellings (Chafe 1980) from neurologically healthy Russian speakers and people with different forms of aphasia and right hemisphere lesions
- Pre-processing was done with ELAN (<https://tla.mpi.nl/tools/tla-tools/elan/>)

Pros:

Strict narrative structure (a unified story, limited entities) helps in analyzing:

- grammatical and lexical features on clause level
- story macrostructure, coherence, interactions, and narrative strategies

EXPERIMENTAL SUBCORPUS

	Healthy subjects	Acoustic-mnesic aphasia	Efferent motor aphasia
Age (mean)	58 (40-84)	48,8 (40-59)	47,8 (30-55)
N of texts	10	5	5
Length (max)	391	419	408
Length (min)	106	195	266
Length (mean)	277,8	292,2	312,8

ACOUSTIC-MNESTIC APHASIA:
1. Fluent (Wernicke's)
2. Loss of connection between meaning and acoustic part
3. Difficulties in understanding directed speech etc.

EFFERENT MOTOR APHASIA:
1. Non-fluent (Broca's)
2. Loss of kinetic speech part
3. Agrammatism etc.

ANNOTATION

COREFERENCE ANNOTATION

Coreference annotation is selecting all the language units that are used to refer to a specific entity.

...мальчик остановился и [pro_synt] взял грушу из корзины. Но потом он понял, что одной ему недостаточно

Language units:

- Noun phrases + bare nouns
- Anaphoric pronouns
- Zero pronouns
- etc.

TOKEN	ENTITY	NP_TYPE	FIRST/REP	EPISODE	#
он	boy	ana	rep	m	2
знал				m	2
что				m	2
здесь				m	2
сборщик	man	NP	first	m	2
груш	pears		rep	m	2
проходил				M	3
человек	m_goat	n	first	M	3
с				M	3
козой	goat	n	first	M	3
и				m	3
не				m	3
взял	m_goat	pro_synt	rep	m	3
ни				m	3
одной				m	3
груши	pears	NP	first	m	3

EPISODE ANNOTATION

An episode is a narrative macrostructure unit.

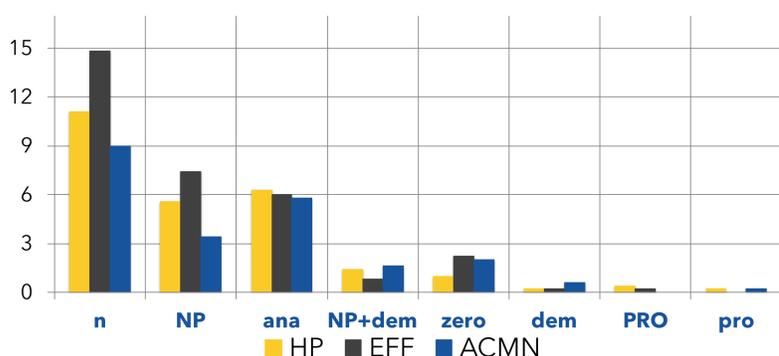
Episode boundaries are based on change of perspective.

This change of perspective can be based on:

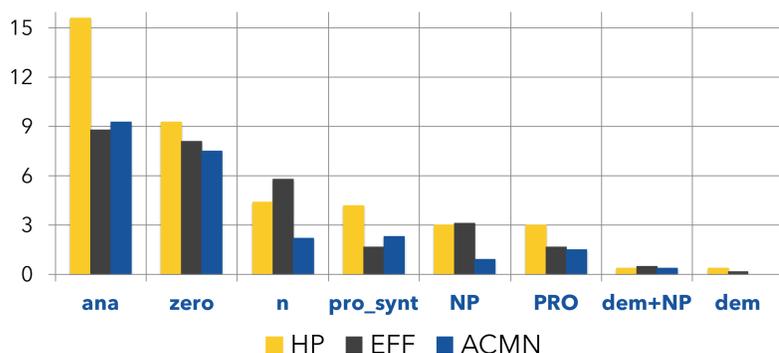
- changing the subject or theme
- transferring from one narrative component to another
 - «static» components: *Beginning, Coda*;
 - «erratic»: *Main line, Interaction, Description*
- switching from Retelling world to the World of the tale

RESULTS

Mean use of every type of referential units as the first entity reference



Mean use of every type of referential units as the repeated entity reference



Referential units as first and repeated reference

As the first reference of a new entity all subjects mostly use bare nouns and noun phrases. The most used NP is a bare (no modifier) noun.

As repeated reference anaphoric pronouns and non-syntactic zeroes are mostly used.

Differences in referencing between groups

We have calculated statistical difference between:

- Mean use of every referential unit in the retellings produced by people from different groups
 - Use of every referential unit normalized by the mean length of the text
- We did not find statistically significant differences in both due to little data. But: the differences that we have identified are important in analysing narrative structure of oral discourse in people with aphasia.

WHAT'S NEXT

Power test

Power test: to reach significance we need 25 more texts in all three groups.

Альтернативные методы анализа

Possible prospects:

- Referential distance
- Referential choice in referring to multiple entities