# Explanation of tags used in the *HereAndNowDatabase.xlsx*

Read this document, which explains the tags used in the database, before you start working with the database *HereAndNowDatabase.xlsx*.

The database contains all the data used in Tore Nesset, Anna Endresen, Laura A. Janda, Anastasia Makarova, Francis Steen and Mark Turner (2013): How ‘here’ and ‘now’ in Russian and English establish joint attention in TV news broadcasts. *Russian Linguistics* 2013.

The file contains six spreadsheets, one for each of the five deictic words analyzed in the article, and one for audiovisual data. The data in the first five spreadsheets are from the Russian National Corpus (RNC, [www.ruscorpora.ru](http://www.ruscorpora.ru), corpus searches performed in January–March, 2012). The data in the last spreadsheet are excerpted from the NewsScape Library of International Television News, a unit of the University of California Library. It is a searchable but not yet public online database ([http://tvnews.library.ucla.edu](http://tvnews.library.ucla.edu/), authentication required).

The spreadsheets include two types of information:

* Information obtained automatically from the RNC when the data was extracted. This data is available for all five words.
* Additional information the words were tagged for. The words were tagged for different properties, as explained in the article and detailed below.

The information from the RNC includes (http://ruscorpora.ru/en/corpora-parameter.html):

1. Left context: the context preceding the deictic word
2. Center: the deictic word
3. Punct: punctuation after the deictic word (if any)
4. Right context: the context following the deictic word
5. Title: the name of the file the example was found in
6. Passport of the examples:
   1. Author: author of the text
   2. Birthday: the year the author of the text was born
   3. Header: name of the text
   4. Created: the date the text was created
7. Additional information about the type of the text:
   1. Sphere
   2. Type
   3. Topic
   4. Publication
   5. Publ\_year
   6. Medium
8. Ambiguity: contains information on whether the text has been disambiguated morphologically and syntactically.

Lists of additional tags for the five deictic words and audiovisual material with short explanations:

**Sejčas**:

1. Temporal reference (these tags are based on Mel’čuk 1985 and explained in section 4 of the article):
   1. S1
   2. S2
   3. S3
   4. S4
2. Type of S1 (this is a further classification of Mel’čuk’s type, discussed in section 4 of the article):
   1. S1a Now (this refers to events happing “right now”, i.e. coinciding with the moment of speech)
   2. S1b Nowadays (this refers to events with a somewhat broader temporal reference, where we are dealing with events that take place “nowadays”.)
   3. not applicable (for other types of NOW than S1)
3. Contrast (these tags show whether the deictic words contrast two events or not, as discussed in section 5 of the article:
   1. No contrast
   2. Temporal
   3. Modal
   4. Generic-Specific
   5. Discourse
4. Aspect 8aspect of the verb in the same clause as the deictic word9
   1. PF (perfective)
   2. IMPF (imperfective)
   3. Not applicable (no verb in the clause)
5. Tense: tense of the verb in the same clause as the deictic word
   1. \_: no verb, impossible to reconstruct a verb
   2. (Pres): copula in present tense
   3. Fut: future tense
   4. Imperat: imperative
   5. Inf: infinitive
   6. Past: past tense of finite verb
   7. Past part: past participle
   8. Pres: present tense of finite verb
   9. Pres Part: present participle
   10. Subj: subjunctive
6. Full context: Full sentence the deictic word was used in
7. Larger context: preceding: for sentences where the contrast was not within the sentence, but in the larger context
8. Larger context: following: for sentences where the contrast was not within the sentence, but in the larger context

**Teper’**:

1. Temporal reference
2. Type of S1
3. Contrast
4. Subtype of temporal contrast:
   1. Past-Present
   2. Present-Future
   3. Past-Future
   4. Not applicable: for cases of no contrast
5. Aspect in the contrasted clause:
   1. PF
   2. IMPF
   3. Not applicable: no verb in the clause
6. Tense in the contrasted clause
7. Aspect of the verb in the same clause
8. Tense in the same clause
9. Full context

**Zdes’**:

1. Type of use (explained in section 3 of the article):
   1. Space
   2. Discourse
   3. Situation
   4. Other
2. Aspect: aspect of the verb in the same clause as the deictic verb
3. Tense: tense of the verb in the same clause as the deictic verb
4. Full context

**Tut**:

1. Type of use (explained in section 3 of the article):
   1. Time
   2. Space
   3. Discourse
   4. Situation
   5. Other
2. Temporal reference
3. Type of contrast
4. Tut zhe:
   1. 1: when the deictic word tut was used with the particle *že*
   2. 0: when the deictic word tut was used without the particle *že*
5. Tense
6. Aspect
7. Full context

**Vot**:

1. Type of use (explained in section 6 of the article):
   1. Time
   2. Space
   3. Discourse
   4. Deictic situation
   5. Other
2. Aspect
3. Tense
4. Full context

**Audiovisual material**:

1. Deictic word:
   1. Sejčas
   2. Teper’
   3. Zdes’
   4. Tut
   5. Vot
2. Full context: the context the deictic word was used in
3. Speaker: who uttered the deictic word:
   1. Anchor in the studio
   2. Reporter in the field
   3. Interviewee
4. Link to video: name of the tagged video file
5. Time in the video: “time stamp” describing when in the video the relevant deictic word was used