In addition, it organizes graphonological features difficulties of previously used feature vectors or unstructured feature sets

As a result, the measure. which uses feature vectors of pronounced sounds in phonological transcription of compared words

pronounciation p honemes proposed for measuring dialectological distance

evaluates in a task-based environment: as a measure of success in automated discovery

which uses the performance-based measure of using the metric in an application

but the central

* what is the optimal way of evaluating :

we evaluate in application…

how to evaluate the quality of different metric settings;

// the advantage is

Phonological distance is used for modelling pronunciation variation across different languages and dialects, even though difficulties have been identified of how to measure edits on individual features and calculate the difference between feature representations associated with compared phonemes.

// another camp – not talking :: linguistic extension of Levenshtein’s metric for dialectological variation

// transcription is used

// synthesis exploit possible synergies: CL applications & linguistically-infomred features and representations… >> for the task of vocabulary building for a hybrid MT system…

discovering translation equivalents for extending coverage of MT systems

Text2

Application of automated cognate identification for hybrid MT tasks

(ACCURAT;

structures have potentially different size, for computing the edit score we use the following procedure:

Precision, Recall, F-measure

we need to address the issue the nature of the

became the standard pronounciation back in standard Russian in most cases is not reflected in the orthography

// bez

// hierarchy

// symmertric

Automated identification of cognates is used for the following two purposes:

1.

We use automated cognate identification

// in first experiments we noted …

// limitations:

Text

application

1. orthography – not phonological transcription
2. principles of orthography
3. use of linguistic features

Text2

Algorithm

appli

Compare b ~ v

Compare t ~ n

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

Annotation scheme: ranking

Practical session for automated annotation of cognates

resulted in sound charts produced by the International Phonetic Association (IPA), where

Agreement on the set of features IPA

>> release of the systems.

crucial for differentiating phonemes.

Some phonemes are distinguished by Distinctive phonological features

and/or a way of irarticulation in the way how phonemes are

correspond to the way how

is characterised by

Distinctive vs. integral features …

or phonemes, or minimal psegments of a language is characterised by

Pronounciation patterns

of differential features of sounds minimal pairs

IPA background

Then the distance is calculated for rewriting of these feature representations rather than rewriting the whole character

principal differences in the phonological system of a language,

use optimal way of representing the system of phonological features

The reason for this can be explained by checking edit matrices and feature vectors of the compared words shown in Figure 1 and Figure 2.

It can be seen from the figures that

and Table 3

r(р) o(о) b(б) i(і) t(т) n(н) y(и) k(к)

0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0

r(р) 1.0 0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0

o(о) 2.0 1.0 0.0 1.0 2.0 3.0 4.0 5.0 6.0

v(в) 3.0 2.0 1.0 1.0 2.0 3.0 4.0 5.0 6.0

e(е) 4.0 3.0 2.0 2.0 2.0 3.0 4.0 5.0 6.0

s(с) 5.0 4.0 3.0 3.0 3.0 3.0 4.0 5.0 6.0

n(н) 6.0 5.0 4.0 4.0 4.0 4.0 3.0 4.0 5.0

i(и) 7.0 6.0 5.0 5.0 5.0 5.0 4.0 3.0 4.0

k(к) 8.0 7.0 6.0 6.0 6.0 6.0 5.0 4.0 3.0

Figure 1 Baseline Levenshtein: Edit distance matrix for

*robitnyk* (робітник) ‘worker’(uk) & *rovesnik* (ровесник) ‘age-mate, of the same age’ (ru)

r(р) o(о) b(б) i(і) t(т) n(н) y(и) k(к)

0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0 8.0

r(р) 1.0 0.0 1.0 2.0 3.0 4.0 5.0 6.0 7.0

o(о) 2.0 1.0 0.0 1.0 2.0 3.0 4.0 5.0 6.0

v(в) 3.0 2.0 1.0 0.8 1.8 2.8 3.8 4.8 5.8

e(е) 4.0 3.0 2.0 1.8 1.2 2.2 3.2 4.2 5.2

s(с) 5.0 4.0 3.0 2.8 2.2 2.0 3.0 4.0 5.0

n(н) 6.0 5.0 4.0 3.8 3.2 3.0 2.0 3.0 4.0

i(и) 7.0 6.0 5.0 4.8 3.8 4.0 3.0 2.2 3.2

k(к) 8.0 7.0 6.0 5.8 4.8 4.6 4.0 3.2 2.2

Figure 3. GPhFeatLev Levenshtein: Edit distance matrix with *hierarchical features* for

*robitnyk* (робітник) ‘worker’(uk) & *rovesnik* (ровесник) ‘age-mate, of the same age’ (ru)

|  |  |
| --- | --- |
| r (р) | ['type:consonant', 'voice:sonorant', 'maner:thrill', 'active:fronttongue', 'passive:palatal'] |
| a (а) | ['type:vowel', 'backness:back', 'height:open', 'roundedness:unrounded', 'palate:nonpalatalizing'] |
| b (б) | ['type:consonant', 'voice:voiced', 'maner:plosive', 'active:labial', 'passive:bilabial'] |
| o (о) | ['type:vowel', 'backness:back', 'height:mid', 'roundedness:rounded', 'palate:nonpalatalizing'] |
| t (т) | ['type:consonant', 'voice:unvoiced', 'maner:plosive', 'active:fronttongue', 'passive:alveolar'] |
| n (н) | ['type:consonant', 'voice:sonorant', 'maner:nasal', 'active:fronttongue', 'passive:alveolar'] |
| i (и) | ['type:vowel', 'backness:front', 'height:close', 'roundedness:unrounded', 'palate:nonpalatalizing'] |
| k (к) | ['type:consonant', 'voice:unvoiced', 'maner:plosive', 'active:backtongue', 'passive:velar'] |

Table 4: Phonological feature vectors for Russian word ‘rabotnik (работник) – ‘worker’