



ASG 1 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

AGTT

Time Series Analysis of Multi-word Units in Late Latin

Maria Sukhareva

Text Technology Group, Goethe-Universität Frankfurt am Main

July 4, 2011



Problem Statement

History, captured by multi-word expressions

ASG 2 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

“Language does not exist apart from culture, that is, from the socially inherited assemblage of practices and beliefs that determines the texture of our lives.”

Edward Sapir

■ Latin word “Virtus” *virtue*



Problem Statement

History, captured by multi-word expressions

ASG 2 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

“Language does not exist apart from culture, that is, from the socially inherited assemblage of practices and beliefs that determines the texture of our lives.”

Edward Sapir

- Latin word “Virtus” *virtue*
- Before Christianity: it was used to describe the virtue of good deeds;



Problem Statement

History, captured by multi-word expressions

ASG 2 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

“Language does not exist apart from culture, that is, from the socially inherited assemblage of practices and beliefs that determines the texture of our lives.”

Edward Sapir

- Latin word “Virtus” *virtue*
- Before Christianity: it was used to describe the virtue of good deeds;
- As Christianity was introduced: it was used to describe Christian moral values;[2]



Problem Statement

What are we looking for?

ASG 3 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

"You shall know a word by the company it keeps. "

J. R. Firth

Multi-word units

units of words which co-occur together more often than expected by chance.



Problem Statement

What are we looking for?

ASG 3 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

"You shall know a word by the company it keeps. "

J. R. Firth

Multi-word units

units of words which co-occur together more often than expected by chance.

- **Graphemic changes:** impossible as we deal with editions;



Problem Statement

What are we looking for?

ASG 3 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

"You shall know a word by the company it keeps. "

J. R. Firth

Multi-word units

units of words which co-occur together more often than expected by chance.

- **Graphemic changes:** impossible as we deal with editions;
- **Morphological changes:** non-native speakers used Latin, therefore, morphology is expected to be stable and show traces of the native language of the author;



Problem Statement

What are we looking for?

ASG 3 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

"You shall know a word by the company it keeps. "

J. R. Firth

Multi-word units

units of words which co-occur together more often than expected by chance.

- **Graphemic changes:** impossible as we deal with editions;
- **Morphological changes:** non-native speakers used Latin, therefore, morphology is expected to be stable and show traces of the native language of the author;
- **Lexical changes:** give hints about social, political and cultural changes in the society;



Data

ASG 4 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

- 1 Patrologia Latina: 1000 years of Latin works (4 A.C. . . .);[1]
- 2 TEI P5 corpus of PL (8.508 files, 12.7 Gb);[7]
- 3 Corpus includes notes and comments in foreign languages;



Approach

MWU Extractions and Ordering

ASG 5 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography



Approach

MWU Extractions and Ordering

ASG 5 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

- 1 Extract MWU;
 - find the appropriate association measure;
 - extract continuous and discontinuous MWUs;
 - set up an evaluation set;



Approach

MWU Extractions and Ordering

ASG 5 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

- 1** Extract MWU;
 - find the appropriate association measure;
 - extract continuous and discontinuous MWUs;
 - set up an evaluation set;
- 2** Ordering of MWUs variants;
 - develop an algorithm to find variants;
 - find a graphic representation;



MWU Extraction

Finding the proper measure

ASG 6 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

The association measure should:



MWU Extraction

Finding the proper measure

ASG 6 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

The association measure should:

- 1 extract continuous and discontinuous MWUs;



MWU Extraction

Finding the proper measure

ASG 6 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

The association measure should:

- 1 extract continuous and discontinuous MWUs;
- 2 deal with low-frequency words;



MWU Extraction

Finding the proper measure

ASG 6 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

The association measure should:

- 1 extract continuous and discontinuous MWUs;
- 2 deal with low-frequency words;
- 3 deal with high-frequency words;



MWU Extraction

Finding the proper measure

ASG 6 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

The association measure should:

- 1 extract continuous and discontinuous MWUs;
- 2 deal with low-frequency words;
- 3 deal with high-frequency words;
- 4 extract multi-word units;



MWU Extraction

Finding the proper measure

ASG 6 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

The association measure should:

- 1 extract continuous and discontinuous MWUs;
- 2 deal with low-frequency words;
- 3 deal with high-frequency words;
- 4 extract multi-word units;
- 5 extract textual co-occurrences (sentential level);



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

1 Mutual Expectation[5]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

1 Mutual Expectation[5]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

1 Mutual Expectation[5]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

1 Mutual Expectation[5]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

1 Mutual Expectation[5]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

1 Mutual Expectation[5]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

- 1 Mutual Expectation[5]
- 2 Normalized MI[4]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

- 1 Mutual Expectation[5]
- 2 Normalized MI[4]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

- 1 Mutual Expectation[5]
- 2 Normalized MI[4]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

- 1 Mutual Expectation[5]
- 2 Normalized MI[4]
- 3 Log-Likelihood[6]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

- 1 Mutual Expectation[5]
- 2 Normalized MI[4]
- 3 Log-Likelihood[6]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Combining Measures

ASG 7 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Fuzzy Decision Making[3]

$A(x) = \sum_{r=1}^n w_r(A_r(x))$, where A is a fuzzy set,

r^{th} is an expert and $x \in X$ is a word combination and $\sum_{r=1}^n w_r = 1$

Experts (r):

- 1 Mutual Expectation[5]
- 2 Normalized MI[4]
- 3 Log-Likelihood[6]

Requirements:

- extract continuous and discontinuous MWUs
- deal with low-frequency words
- deal with high-frequency words
- extract longer multi-word units
- extract textual co-occurrences (sentential level)



MWU Extraction

Mutual Expectation

ASG 8 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Counting the fair point of expectation

Expectation of	Knowing the gapped 3-gram
Republic	[- +1 of +2 Yugoslavia]
of	[Republic +1 - +2 Yugoslavia]
Yugoslavia	[Republic +1 of +2 -]

Mutual Expectation:

$$ME([w_1 \dots d_{1i} w_i \dots d_{1n} w_n]) =$$

$$f([w_1 \dots d_{1i} w_i \dots d_{1n} w_n]) \times NE([w_1 \dots d_{1i} w_i \dots d_{1n} w_n])$$

d_{ij} is the distance between w_j and w_i and NE (normalized expectation) is the ratio of probability of occurrence of the n-gram to FPE.



MWU Extraction

Normalized MI and Log-Likelihood

ASG 9 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Normalized MI:

$Normalized_MI([w_1 \dots d_{1i} w_i \dots d_{1n} w_n]) =$

$$\log_2 \frac{N \times f([w_1 \dots d_{1i} w_i \dots d_{1n} w_n])}{FPD_{MI}([w_1 \dots d_{1i} w_i \dots d_{1n} w_n])}$$

Log-Likelihood:

$$LogLikelihood = 2 \sum_{ij} O_{ij} \log \frac{O_{ij}}{E_{ij}}$$

where O_{ij} is the observed frequency and E_{ij} is the expected frequency and i and j are rows and column indices in the contingency table



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Variants Typology

Variants

Surface

Graphemic Morphological Lexical Syntactic Semantic

SurfMorph SemMorph SurfLex SemLex SemSynt SurfSynt



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Variants Typology

Variants



Surface

Graphemic Morphological Lexical Syntactic Semantic

SurfMorph SemMorph SurfLex SemLex SemSynt SurfSynt



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

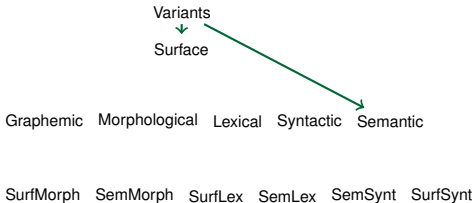
Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Variants Typology





Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

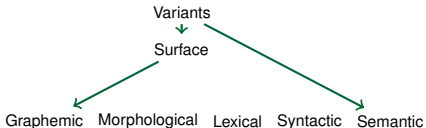
Evaluation

Conclusion

Bibliography

Variants Typology

1 **Graphemic:** interchange of letters 'u' and 'v';



SurfMorph SemMorph SurfLex SemLex SemSynt SurfSynt



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

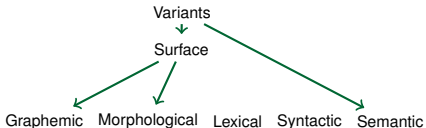
Evaluation

Conclusion

Bibliography

Variants Typology

1 **Graphemic:** interchange of letters 'u' and 'v';



SurfMorph MorphSem SurfLex LexSem SemSynt SurfSynt



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

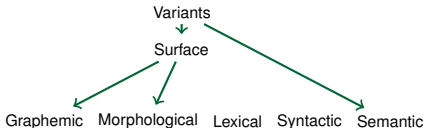
Evaluation

Conclusion

Bibliography

Variants Typology

1 **Graphemic:** interchange of letters 'u' and 'v';



SurfMorph MorphSem SurfLex LexSem SemSynt SurfSynt



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

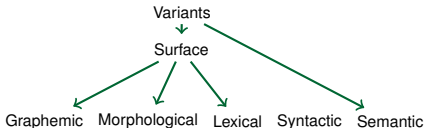
Evaluation

Conclusion

Bibliography

Variants Typology

1 **Graphemic:** interchange of letters 'u' and 'v';



SurfMorph MorphSem SurfLex LexSem SemSynt SurfSynt



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

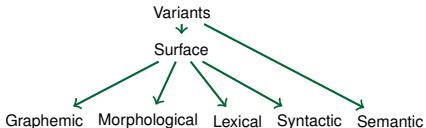
Evaluation

Conclusion

Bibliography

Variants Typology

1 **Graphemic:** interchange of letters 'u' and 'v';



SurfMorph MorphSem SurfLex LexSem SemSynt SurfSynt



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

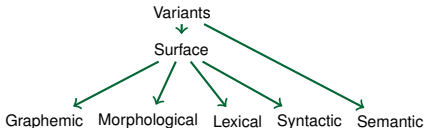
Evaluation

Conclusion

Bibliography

Variants Typology

1 **Graphemic:** interchange of letters 'u' and 'v';



SurfMorph MorphSem SurfLex LexSem SemSynt SurfSynt



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

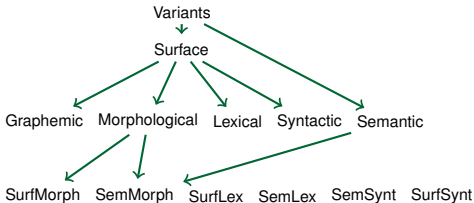
Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Variants Typology



1 Graphemic: interchange of letters 'u' and 'v';

2 SemMorph: verbal suffix /sk/ which once had a meaning of a chaotic action, totally lost this meaning;

3 SurfMorph: any reduced inflections, for example;



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

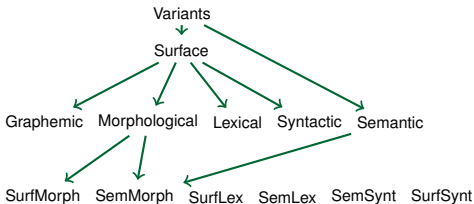
Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Variants Typology



1 **Graphemic:** interchange of letters 'u' and 'v';

2 **SemMorph:** verbal suffix /sk/ which once had a meaning of a chaotic action, totally lost this meaning;

3 **SurfMorph:** any reduced inflections, for example;



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

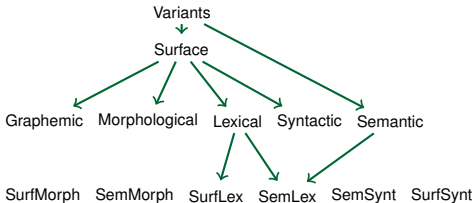
Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Variants Typology



1 **Graphemic:** interchange of letters 'u' and 'v';

2 **SemMorph:** verbal suffix /sk/ which once had a meaning of a chaotic action, totally lost this meaning;

3 **SurfMorph:** any reduced inflections, for example;

4 **SurfLex:** absolute synonyms;

5 **SemLex:** as in example with 'virtus';



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

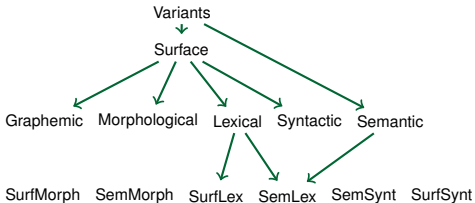
Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Variants Typology



- 1 Graphemic:** interchange of letters 'u' and 'v';
- 2 SemMorph:** verbal suffix /sk/ which once had a meaning of a chaotic action, totally lost this meaning;
- 3 SurfMorph:** any reduced inflections, for example;
- 4 SurfLex:** absolute synonyms;
- 5 SemLex:** as in example with 'virtus';



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

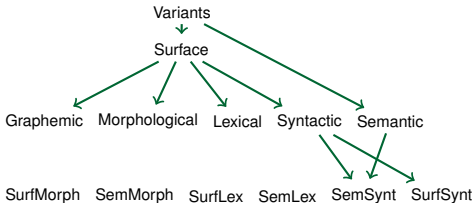
Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Variants Typology



- 1 Graphemic:** interchange of letters 'u' and 'v';
- 2 SemMorph:** verbal suffix /sk/ which once had a meaning of a chaotic action, totally lost this meaning;
- 3 SurfMorph:** any reduced inflections, for example;
- 4 SurfLex:** absolute synonyms;
- 5 SemLex:** as in example with 'virtus';
- 6 SemSynt:** loss of idiomaticity;
- 7 SurfSynt:** the same meaning of the phrase after changed syntax;



Ordering of MWU Variants

Variants Typology

ASG 10 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

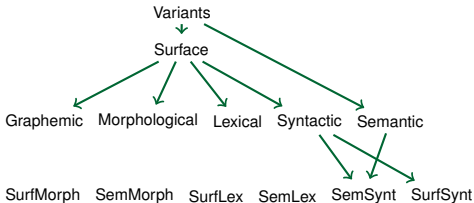
Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Variants Typology



- 1 Graphemic:** interchange of letters 'u' and 'v';
- 2 SemMorph:** verbal suffix /sk/ which once had a meaning of a chaotic action, totally lost this meaning;
- 3 SurfMorph:** any reduced inflections, for example;
- 4 SurfLex:** absolute synonyms;
- 5 SemLex:** as in example with 'virtus';
- 6 SemSynt:** loss of idiomaticity;
- 7 SurfSynt:** the same meaning of the phrase after changed syntax;



Ordering of MWU Variants

Creating Minimum Spanning Trees

ASG 11 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

- 1 Decision making algorithm of Kacprzyk and Nurmi;[8]
- 2 Five experts: lexical, morphological, graphemic, syntactic and contextual;
- 3 Creation of a fuzzy set of the MWUs;



Ordering of MWU Variants

Expert Example

ASG 12 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Morphological Expert

$$\mu^k = \begin{cases} 1.0 & \text{if the word-form is the same} \\ 0.75 & \text{if lemma is the same} \\ 0.5 & \text{if lemma is different but stem is the same} \\ 0.25 & \text{if stem and lemma are different but not inflectional morphology} \\ 0.0 & \text{all is different} \end{cases}$$

(grades are rough and will be learnt from the data)

$$C_Q = \{(A1, v_Q^1), (A2, v_Q^2), (A3, v_Q^3) \dots (A1, v_Q^n)\}. \quad (2)$$

Minimum spanning tree edges are inverted: $1 - v_Q^n$



Ordering of MWU Variants

Minimum Spanning Trees

ASG 13 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

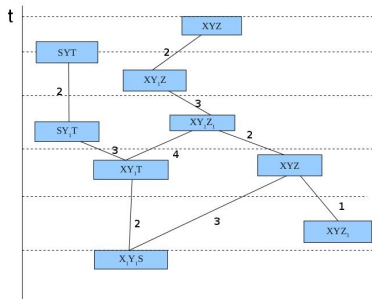
Conclusion

Bibliography

Minimum Spanning Trees

The path with the smallest sum has the highest probability that the MWUs are variants of each other

A Sample Graph of 3-grams for Counting Minimum Spanning Trees





Evaluation

Pain in the Neck

ASG 14 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

Problem:

- no annotated data
- impossible to annotate it manually.

Solution:

- extract all the possible MWUs
- create a set of possible variants
- manually annotate true positives
- create a gold standard of around 1000 MWU variants.



Conclusion

ASG 15 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography

- MWUs can give clues to historians about the social, political and economical life.
- Two tasks: extract MWUs and find their variants
- Final representation should have: time and membership values



ASG 16 16

Maria Sukhareva

Introduction

Data

Approach

MWU Extraction

Ordering of MWU Variants

Evaluation

Conclusion

Bibliography



Patrologia latina, <http://en.wikipedia.org/wiki/PatrologiaLatin>.



Virtue, <http://en.wikipedia.org/wiki/Virtue>.



Raj Kishor Bisht and H.S. Dhama, Fuzzy set theoretic approach to collocation extracion, International Journal of Computer Applications 5 (2010), no. 3.



Gael Dias, Sylvie Guillore, and Jose Gabriel Pereira Lopes, Normalization of association measures for multiword lexical unit extraction, International Conference on Artificial and Computational Intelligence for Decision, Control and Automation in Engineering and Industrial Applications, 2000.



Gael Dias, Jose Gabriel Pereira Lopes, and Sylvie Guillore, Mutual expectation: A measure for multiword lexical unit extraction, VEXTAL'99, November 1999.



Stefan Evert, Corpora and collocations, Corpus Linguistics. An International Handbook (2007).



Alexander Mehler, Nils Diewald, Ulli Waltinger, Rüdiger Gleim, Dietmar Esch, Barbara Job, Thomas Küchelmann, Olga Pustynnikov, and Philippe Blanchard, Evolution of Romance language in written communication: Network analysis of late Latin and early Romance corpora, Leonardo 44 (2011), no. 3.



Predgrad Prodanovic, Fuzzy set ranking methods and multiple expert decision making, Tech. report, Water Resources Research Report, 2001.