

T. Erjavec, J. Gros (ur.)

# JEZIKOVNE TEHNOLOGIJE ZA SLOVENSKI JEZIK

Zbornik konference 'Jezikovne tehnologije za slovenski jezik'  
6-7 oktober 1998, Ljubljana

## PREDGOVOR

V pričujočem zborniku so objavljeni prispevki s konference "Jezikovne tehnologije za slovenski jezik", ki je potekala 6. in 7. oktobra 1998 v Ljubljani, v okviru multikonference "Informacijska družba", IS'98.

Informacijska družba se konstituira predvsem skozi izmenjavo jezikovnih podatkov; jezikovne tehnologije (JT) zajemajo metode in načine, ki omogočajo računalniško izmenjavo, strukturiranje, pregledovanje, hranjenje in prevajanje jezikovnih podatkov. Konferenca je bila posvečena uporabi in razvoju JT za slovenski jezik, njen namen pa je bil predstaviti obstoječe JT pri nas in v svetu, možnosti njihove uporabe in smernice prihodnjega razvoja na tem področju.

Zbornik vsebuje 26 prispevkov, ki obravnavajo govorne tehnologije in glasoslovje, uporabo računalnikov v prevajanju in poučevanju jezikov, računalniške korpuse, standarde zapisa jezikovnih podatkov in iskanje informacij po internetu. Članki podajajo raziskovalne in praktične rezultate, pa tudi preglede in zamisli za razvoj tega področja v prihodnje. Med pregledi najdemo tako predstavitev delovanja Evropske unije na področju jezikovnih tehnologij kot tudi predstavitev delovanja slovenskih vladnih in izobraževalnih institucij, podjetij ter ljubiteljev. Organizatorji bi se radi zahvalili vsem, ki so prispevali k uspehu konference: vabljenim predavateljem in avtorjem prispevkov, programskemu odboru za recenzentsko delo ter organizatorjem IS'98.

## PREFACE

These proceedings contain the contributions to the conference "Language Technologies for the Slovene Language", which took place on the 6th and 7th October in Ljubljana, in the scope of the Information Society Multiconference, IS'98.

The Information Society is constituted primarily through the exchange of language data; Language Technologies (LT) comprise computational tools and methods that assist in authoring, exchange, structuring, searching, storage and translation of language data. The focus of the conference was on the usage and development of LT for the Slovene language. It aimed to introduce LT and their use, give an overview of the state of the art in Slovenia and abroad, and discuss plans for the future of Slovene LT.

These Proceedings contain 26 contributions, which discuss speech technologies and phonology, the use of computers in translating and language learning, computer corpora, standardisation of language data encoding, and information retrieval from the Internet. The papers present research and applied results as well as give overviews and ideas for the future of the field. The overviews contain a synopsis of EU activities in the field of LT and presentations by Slovene government and educational institutions, companies and enthusiasts.

The organisers would like to thank the many people who contributed to the success of the conference: the invited lecturers and authors of contributions, the Programme Committee of the conference and the Organising Committee of IS'98.

Tomaz Erjavec  
Ljubljana, september 1998.

## Programski odbor / Programme Committee

<i>prof.dr. Vladimir Batagelj</i>	Fakulteta za matematiko in fiziko, Univerza v Ljubljani
<i>dr. Jure Dimec</i>	Institut za biomedicinsko informatiko, Medicinska fakulteta, Univerza v Ljubljani
<i>dr. Tomaž Erjavec</i>	Odsek za inteligentne sisteme, Institut Jožef Stefan
<i>prof.dr. Borka Jerman-Blažič</i>	Laboratorij za odprte sisteme in mreže, Institut Jožef Stefan
<i>prof.dr. Zdravko Kačič</i>	Inštitut za elektroniko, Fakulteta za elektroteh., rač. in inf., Univerza v Mariboru
<i>prof.dr. Irena Kovačič</i>	Oddelek za prevajanje in tolmačenje, Filozofska fakulteta, Univerza v Ljubljani
<i>prof.dr. France Mihelič</i>	Laboratorij za umetno zaznavanje, Fakulteta za elektrotehniko, Univerza v Ljubljani
<i>doc.dr. Vesna Požgaj-Hadži</i>	Oddelek za slovanske jezike in književnosti, Filozofska fakulteta, Univerza v Ljubljani
<i>doc.dr. Marko Stabej</i>	Oddelek za slovanske jezike in književnosti, Filozofska fakulteta, Univerza v Ljubljani

## Organizacijski odbor / Organising Committee

<i>dr. Tomaž Erjavec</i>	Institut Jožef Stefan, (predsednik / chair)
<i>Vojko Gorjanc</i>	Filozofska fakulteta, Univerza v Ljubljani
<i>dr. Jerneja Gros</i>	Fakulteta za elektrotehniko, Univerza v Ljubljani
<i>Anica Rant</i>	Društvo znanstvenih in tehničnih prevajalcev Slovenije

## Organizacijski odbor IS'98 / Organising Committee IS'98

<i>prof. dr. Matjaž Gams</i>	(predsednik / chair)
<i>dr. Tanja Urbančič</i>	(podpredsednica / co-chair)
<i>Tomaž Bojko</i>	<i>dipl. oec.</i>
<i>Damjan Demšar</i>	<i>dipl. ing.</i>
<i>mag. Matija Drobnič</i>	
<i>Špela Korenčan</i>	
<i>mag. Viljem Križman</i>	
<i>mag. Dunja Mladenič</i>	
<i>dr. Matjaž Ostroveršnik</i>	
<i>Mili Remetič</i>	
<i>Radoš Skrt</i>	<i>dipl. oec.</i>
<i>dr. Blaž Zupan</i>	

# Kazalo / Table of Contents

## Vabljena predavanja / Invited Lectures

Language Resources, TELRI and Multilingual Lexical Semantics <i>W. Teubert</i> .....	1
Language Technology and Multilinguality — The European Dimension <i>P. Andersen</i> .....	9
Natural Language Processing at the Xerox Research Centre Europe <i>J. P. Chanod</i> .....	14

## Uvodna predavanja / Introductory Lectures

On the Vowel System in Present-Day Standard Slovene <i>T. S. Rejec</i> .....	22
VERBMOBIL: A Speech-to-Speech Translation System <i>D. Čavar, W. Menzel</i> .....	25
AMEBIS IN JEZIKOVNE TEHNOLOGIJE <i>M. Romih</i> .....	29
GNUsl: PROSTO PROGRAMJE IN SLOVENŠČINA <i>A. Košir, P. Peterlin, T. Erjavec</i> .....	35

## Iskanje informacij / Information Retrieval

ISKALNIK ZA SLOVENSKE IN ANGLEŠKE DOKUMENTE NA SVETOVNEM SPLETU <i>J. Dimec, S. Džeroski, L. Todorovski, D. Hristovski</i> .....	42
PRIPOMOČKI ZA ISKANJE PO POLNEM BESEDILU V SLOVENŠČINI <i>A. Tomažič</i> .....	49

## Govor / Speech

AKUSTIČNA SPEKTRALNA FFT ANALIZA SAMOGLASNIŠKEGA SISTEMA SLOVENSKEGA JEZIKA: FORMANTI SLOVENSКИH SAMOGLASNIKOV <i>M. Ozbič</i> .....	55
SLOVENSKI GOVOR NA INTERNETU <i>T. Šef, A. Dobnikar, M. Gams, M. Grobelnik</i> .....	60
GOVOREČI RAČUNALNIK <i>J. Gros, F. Mihelič, N. Pavešič</i> .....	65
UPORABA GOVORNE TEHNOLOGIJE PRI AVTOMATIZACIJI DALJINSKIH TELEFONSKIH STORITEV <i>Z. Kačič, B. Horvat, B. Imperl, A. Miksič, J. Kaiser, M. Rojc, M. Sepešy-Maučec</i> ...	71
GOVORNA POIZVEDOVANJA V ČEŠČINI, NEMŠČINI, SLOVAŠČINI IN SLOVENŠČINI <i>F. Mihelič, I. Ipšič, J. Gros, K. Pepelnjak, S. Dobrišek, N. Pavešič</i> .....	76

ZUNAJJEZIKOVNE OKOLIŠČINE NEIDEALNEGA GOVORA <i>P. Vitez</i> .....	81
---	----

## **Prevajanje / Translation**

RAČUNALNIKI V PREVAJANJU: KAKO RAZVRSTITI PREDNOSTI? <i>J. Lajovic</i> .....	84
PROGRAMI S POMNILNIKOM PREVODOV S STALIŠČA MOREBITNEGA UPORABNIKA <i>Š. Vintar</i> .....	87
PROBLEMATIKA PREVAJANJA ZAKONODAJE EVROPSKE UNIJE <i>A. Krstič</i> .....	92
VIRTUALNA UČILNICA: UPORABA INTERNETA PRI POUČEVANJU TUJIH JEZIKOV <i>A. Pisanski</i> .....	96

## **Korpusi in standardizacija / Corpora and Standardisation**

IZGRADNJA INFRASTRUKTURE POTREBNE ZA RAZVOJ GOVORNE TEHNOLOGIJE ZA SLOVENSKI JEZIK <i>Z. Kačič, B. Horvat</i> .....	100
GOPOLIS: SLOVENSKA PODATKOVNA ZBIRKA GOVORJENIH POIZVEDOVANJ <i>S. Dobrišek, J. Gros, I. Ipšič, K. Pepelnjak, F. Mihelič, N. Pavešič</i> .....	105
GOVORJENA BESEDILA IN KORPUS SLOVENSKEGA JEZIKA <i>S. Kranjc</i> .....	109
KORPUSI V PREVODOSLOVJU <i>N. Hirci</i> .....	113
STANDARDIZACIJA ZAPISA JEZIKOVNIH PODATKOV <i>T. Erjavec</i> .....	119
KORPUS FIDA <i>T. Erjavec, V. Gorjanc, M. Stabej</i> .....	124
DIGITALNI ZAPIS SLOVENSКИH ZNAKOV <i>P. Peterlin, A. Košir, T. Erjavec</i> .....	128

Kazalo po avtorjih / Author Index .....	133
---	-----

## Kazalo po avtorjih / Author Index

Andersen, Poul .....	9
Chanod, Jean Paul .....	14
Ćavar, Damir .....	25
Džeroski, Sašo .....	42
Dimec, Jure .....	42
Dobnikar, Aleš .....	60
Dobrišek, Simon .....	76, 105
Erjavec, Tomaž .....	35, 119, 124, 128
Gams, Matjaž .....	60
Gorjanc, Vojko .....	124
Grobelnik, Marko .....	60
Gros, Jerneja .....	65, 76, 105
Hirci, Nataša .....	113
Horvat, Bogomir .....	71, 100
Hristovski, Dimitar .....	42
Imperl, Bojan .....	71
Ipšič, Ivo .....	76, 105
Kačič, Zdravko .....	71, 100
Kaiser, Janez .....	71
Košir, Aleš .....	35, 128
Kranjc, Simona .....	109
Krstič, Adriana .....	92
Lajovic, Jaro .....	84
Menzel, Wolfgang .....	25
Mihelič, France .....	65, 76, 105
Miksič, Andrej .....	71
Ozbič, Martina .....	55
Pavešič, Nikola .....	65, 76, 105
Pepelnjak, Karmen .....	76, 105
Peterlin, Primož .....	35, 128
Pisanski, Agnes .....	96
Rejec, Tatjana Srebot .....	22
Rojc, Matej .....	71
Romih, Miro .....	29
Sepešy-Maučec, Mirijam .....	71
Stabej, Marko .....	124
Šef, Tomaž .....	60
Teubert, Wolfgang .....	1
Todorovski, Ljupčo .....	42
Tomazič, Anton .....	49
Vintar, Špela .....	87
Vitez, Primož .....	81