

**SLOVENIAN DATABASE SNABI SFN (Slovenian Fixed Network)
LARGE TELEPHONE SPEECH CORPUS**

Author(s):	Zdravko Kačič, Bogomir Horvat, Aleksandra Zögling Markuš, Robert Veronik, Matej Rojc, Andrej Žgank, Mirjam Sepesy Maučec, Tomaž Rotovnik
Institute:	University of Maribor, Faculty of Electrical Engineering and Computer Science, Centre for Language Technologies
Address:	Smetanova 17, 2000 Maribor, Slovenia
Email:	kacic@uni-mb.si, bogo.horvat@uni-mb.si, sandra.zogling@uni-mb.si, robert.veronik@uni-mb.si, matej.rojc@uni-mb.si, andrej.zgank@uni-mb.si, mirjam.sepesy@uni-mb.si, tomaz.rotovnik@uni-mb.si
Date:	February 2002 (slightly modified for CLARIN.SI distribution, February 2016)
Version:	1.2

Contents

1. INTRODUCTION	4
1.1 Speech file format	5
1.2 File nomenclature	5
1.3 Directory structure	6
1.4 Label files	7
1.4.1 Label file header	7
1.4.2 Label file body	10
2. DATABASE DESIGN AND COLLECTION	12
2.1 Recording site and platform	12
2.2 Speaker recruitment	12
2.3 Design of prompting and prompt-sheet	13
2.4 Transcription	13
3. DATABASE CONTENTS DEFINITION	13
3.1 Alphabet	17
3.2 Passage - phonetically rich sentences	17
3.3 Words	18
3.4 Numbers and number sets	20
3.5 MMC corpus	21
3.6 Lingua	22
4. SPEAKER DEMOGRAPHIC INFORMATION	23
4.1 Accent/Regions	23
4.2 Speaker characteristics	24
5. THE LEXICON	25
6. RECORDING CONDITIONS	26
6.1 Environments	26
6.2 Network	26
6.3 Handsets	26
7. TEST SET DEFINITION	26

7.1	Selection of test speakers	27
7.2	Test Set Specifications	27
7.2.1	Test session list file (*.SES)	27
7.2.2	Test description file (*.DSC)	28
7.3	Tests defined for SNABI SFN database	30
APPENDIX A		32
8. SET OF USED SENTENCES FOR CORPUS MMC		32
8.1	Corpus MMC 1A	32
8.2	Corpus MMC 1B	33
8.3	Corpus MMC 2A	35
8.4	Corpus MMC 2B	36
8.5	Corpus MMC 3A	38
8.6	Corpus MMC 3B	39
9. SET OF USED SENTENCES FOR CORPUS LINGUA		40
9.1	Corpus Lingua 1A	41
9.2	Corpus Lingua 1B	42
9.3	Corpus Lingua 2A	45
9.4	Corpus Lingua 2B	47
9.5	Corpus Lingua 3A	49
9.6	Corpus Lingua 3B	51
9.7	Corpus Lingua 4A	53
9.8	Corpus Lingua 4B	55
9.9	Corpus Lingua Passage	57
10. SET OF USED WORDS, NUMBERS AND SET OF DIGITS FOR CORPUS WORDS		59
10.1	Corpus WRD	59
10.2	Corpus ABC	60
10.3	Corpus SETS	61
10.4	Corpus NUMBERS	61

1. Introduction

The SNABI Slovenian Database for Fixed Telephone Network was recorded. The database was collected at the University of Maribor (Slovenia).

This database includes telephone recordings of speakers recorded directly over the fixed PSTN using the analogue telephone interface and recorded on DAT tape.

The list of the distribution disks and directories are contained in the README.TXT file, stored in the root directory of each CD-ROM.

File types are identified with the following extensions:

*.DOCX	- MicroSoft Word document
*.DSC	- UTF-8 text file
*.LST	- UTF-8 text file
*.PDF	- Adobe Portable Document Format file
*.RAW	- uncompressed A-law speech signal file
*.TBL	- UTF-8 text file
*.TXT	- UTF-8 text file
*.VCB	- UTF-8 text file
*.WSP	- UTF-8 text file

The files in this distribution have the following directory structure:

/:

COPYRIGHT.TXT	- copyright notice
README.TXT	- describes files in the CD-ROM
SFN/	- data directory

SFN /DOC:

DESIGN.DOC	- Slovenian database documentation file
DESIGN.PDF	- Slovenian database documentation file
SAMPALEX.PDF	- SAMPA table
MIS_SAMP.TXT	- list of missing samples in the database
MAN_SEG.TXT	- list of manually segmented files

SFN /INDEX:

CONTENTS.LST	- file/utterance/speaker index table
F1TSTSI.SES	- list of selected speakers in the first test set
F2TSTSI.SES	- list of selected speakers in the second test set

SFN /TEST:

F101SL.DSC	- test description files
...	
F1S01SL.LST	- test list files
...	
F1V01SL.VCB	- test vocabularies

SFN /TABLE:

LEXICON.TBL	- full lexicon table
SPEAKERF.TBL	- speaker information table

SFN /:

MMC/
LINGUA/
WORDS/

SFN / MMC/:

1A/

1B/

...

SFN / MMC/1A:

spk_00/

-directories for each speaker

spk_01/

...

SFN / MMC/1A /spk_00/:

00MT1002.RAW

- speech signal file

00MT1002.TXT

- SAM label file

00MT1002.WSP

- SAM manually segmented label file

...

1.1 Speech file format

The speech was recorded over the fixed PSTN, using the analogue telephone interface and DAT recorder and then transferred over the digital connection (using DAT-Link, manufactured by Townshend Computer Tools) to workstation, where it was stored with 16 kHz sampling rate and 16 bit linear quantization with MSB-LSB byte order.

1.2 File nomenclature

File names follow the ISO 9660 file name conventions (8 plus 3 characters) according to the main CD-ROM standard. The following template is used:

XX Y V ZZZZ . NL M

where:

XX	denotes speaker ID code
Y	code of the sub-lexicon used (W – words, N – numbers, A – alphabet, S - number string, M – MMC corpus, L – lingua corpus, P - passage)
V	type of recording (T - telephone, S - studio)
ZZZZ	sentence ID number
N	type of pronunciation (W – isolated word, S – sentence)
L	Language
M	file type (S – signal, W – segmentation and labeling on word level, P – segmentation and labeling on phone level)

Table 1 - SNABI filename convention

As it is useful for users to clearly identify the speech file contents by looking at the filename we have specified the corpus code to support one letter corpus identifier and one letter item identifier, by the following table.

Nr.	CORPU	CONTENT
	S	

1	Words	68 words (command words, names of capital cities in Europe)
2	Abc	Alphabet - 26 letters
3	Sets	Number strings (set of 20 six digits long number strings)
4	Numbers	Set of digits from 0 to 9
5	MMC	Set of 529 different sentences - divided into three sub-corpora – MMC 1, MMC 2, MMC 3
6	Lingua	Set of 911 different sentences – divided into four sub-corpora – Lingua 1, Lingua 2, Lingua 3, Lingua 4
7	Passage	90 contextual dependent sentences

Table 2 - Overview of the text corpora used in recording the database

The proposed format uses mnemonic values. It permits selection of all files belonging to one of seven corpora by using one command (e.g. in DOS “`dir /s/b ??????C*`”, in UNIX “`find . -name "??????C*" -print`”).

1.3 Directory structure

The database contains speech and corresponding annotation files that are stored in a specified directory structure. The directory structure is a content dependent directory structure, where files are organized according to the corpora used and further to sub-corpora. At the highest level the structure consists directories: words, lingua, and MMC. At the lowest level there are directories for corresponding sub-corpora (for example, lingua 1, lingua 2, lingua 3, and lingua 4, for the lingua corpus). Each of sub-corpus in Lingua and MMC corpus are divided in two sub-corpora – A and B. Every directory at this level contains the speech material of all the speakers that read the text of the sub-corpus. The directory named /doc contains all the documentation of the database. The documentation consists of the table of the SAMPA symbols used in phonetic transcription of the text, lexicon, speaker information table, and UTF-8 table.

Both signal and label files are put in the same directory.

In addition to the previous structure the following directories are used to store some other files:

/<database>/DOC	documentation files, including subword occurrence files
/<database>/TABLE	speaker and lexicon tables
/<database>/INDEX	index files - contents file

Table 3 - Non-speech related directory structure

The root directory contains three files:

- a “README.TXT” ASCII file describing all files in the CD-ROM; signal and label files are reported by specifying their templates;

- a “DISK.ID” ASCII file containing the volume name (9 characters long); it supplies the volume label to UNIX systems that are unable to read the physical volume label, e.g. “SI_SFN_01”.
- a “COPYRIGHT.TXT” ASCII file to protect the authors rights.

All these support files are duplicated in each CD-ROM.

1.4 Label files

Each signal file has associated label files with ASCII SAM labels.

Associated SAM label files are text files where each row can be up to 80 chars long (according to the DOS format). Rows are produced according to the main SAM paradigm:

ABC: x, y, z, ...

Where:

- ABC is a three letter mnemonic followed by a colon: no spaces are allowed between them, so we can define as SAM-mnemonic the set “ABC:”;
- after the mnemonic all the defined items are separated by commas;
- missing items are accepted and nothing needs to be put between commas to substitute them;
- spaces are not significant but good pretty printing rules allow a better readability.

Software tools should use the C instruction `fgets(buffer, fp, 83)`¹ to read each row and afterwards they should scan the filled buffer looking for the commas: the `scanf()` C function can result in incorrect reading.

A label file begins with the mnemonic “LHD:” and end with “ESG:”. The mnemonic “LBD:” splits a label file into two sections: the LABEL FILE HEADER and the LABEL FILE BODY.

All mnemonics defined are listed below. For each one the explanation, the format and the domain accepted for the related items are given. Mnemonics for which no values exist are provided without any values.

1.4.1 Label file header

In the following all the attributes, grouped by similarity, are presented that have been selected as representative of a speech database.

As mentioned before, the mnemonic related items are separated using commas. Spaces are not significant but good pretty printing rules should allow a better readability. The format defined for the various items is described in the second column of the following tables by using a C `printf()` format string.

The order of the mnemonics in the label file header is irrelevant except that “LHD:” is the first mnemonic and “LBD:” is the last one.

¹ Where 83 = 80 + <CR> + <LF> + ‘\0’ on UNIX systems. 82 is enough on DOS/Windows/OS2 systems, because the couple <CR> <LF> is replaced by the escape char ‘\n’. For software portability it is better to fix the buffer to 83 chars.

Identification rows

The starting mnemonic “LHD:” identifies the format name (i.e. “SAM”) and the version number (i.e. “5.10”).

“ESG:” terminates a label file and it provides a check for accidental file truncation.

mnem.	item format	example	comments
LHD :	“%s, %d.%02d”	SAM, 5.10	format name + version
ESG :	“”		end of label file

“CMT:” identifies a comment row.

mnem.	item format	example	comments
CMT :	“%.75s”	This is a comment row	comment row

Session rows

“DBN:” database name.

“VOL:” volume identification name. It coincides with the volume name of the mastered CD-ROM. The volume sequence number starts from “1”.

mnem.	item format	example	comments
DBN :	“%.75s”	SNABI_telephone	database name
VOL :	“%.11s”	SI_SFN_01	database volume ID

Recording condition rows

“REG:” calling region.

“ENV:” calling environment,

“NET:” telephone network,

“PHM:” hand-set type used.

The set of possible values and the distributions are reported in section 6. All fields are inserted in the label files by using the following format:

mnem.	item format	example	comments
REG :	“%.1d”	1	calling region
ENV :	“%.75s”	OFFICE	calling environment
NET :	“%.75s”	PSTN	telephone network
PHM :	“%.75s”	TOUCHTONE	telephone hand set model

Speaker rows

“SCD:” unique speaker code;

“SEX:” speaker gender;

“AGE:” speaker age;

“ACC:” speaker accent, i.e. the regional/dialectical colouring factor;

The set of possible values and the distributions are reported in section 4. All fields are inserted in the label files by using the following format:

mnem.	item format	example	comments
SCD :	"%2c"	CU	speaker code
SEX :	"%c"	M	speaker gender
AGE :	"%d"	53	speaker age
ACC :	"%.s"	1	speaker accent

File rows

"DIR:" directory of the signal file. It is reported in DOS format with a leading backslash (it allows unambiguous paths that start from the root directory).

"SRC:" signal file name.

"CCD:" corpus code. As reported before, this field is included also in the filename construction and has been designed as a alphanumeric character string. (MMC, LINGUA, WORDS)

"CRP:" corpus repetition. It is void i.e. "nil".

"REP:" the place where the recording machine was located.

"RED:" recording date of the signal file.

"BEG:" the begin field specifying the starting point of the speech inside the data file. It is set to zero.

"END:" the end field specifies the size of speech file.

mnem.	item format	example	comments
DIR :	"/%.8s/.../%.8s"	/SFN/MMC/1A/spk_01	signal file directory
SRC :	"%8s.%.3s"	01MT1002.WSS	signal file name
CCD :	"%.s"	MMC	corpus code
CRP :	"%.2d"	-	corpus repetition (SDB)
REP :	"%s, %s, %s"	UMARIB, MARIBOR, SLOVENIA	recording place: place, city, country
RED :	"%02d/%.3s/%4d"	17/Mar/1994	recording date
RET :	"%02d/%.2d/%2d"	-	recording time
BEG :	"%ld"	0	labelled sequence start position
END :	"%d"	File length - 1	labelled sequence end position

Data file coding rows

All information about the coding format are described by using the fields listed below:

"SAM:" the sampling frequency in Hertz (set to 16000);

"SNB:" the number of bytes per each sample (set to 2);

"SBF:" the sample byte order.

"SSB:" number of significant bits per sample (set to 16);

"QNT:" the quantisation attribute specifies the speech coding chosen i.e. "lin").

mnem.	item format	example	comments
SAM:	“%d”	16000	sampling frequency
SNB:	“%d”	2	number of (8-bit) bytes per sample
SBF:	“%75s”	MOTOROLA	sample byte order
SSB:	“%d”	16	number of significant bits per sample
QNT:	“%.75s”	lin	Quantisation

1.4.2 Label file body

The label file body starts with the mnemonic LBD.

“LBD:” that separates it from the label file header; in the label file body are described all the items uttered by a speaker.

Each prompted utterance has been stored in a separate file. We have two different labelling rows that independently store the orthographic prompt and the transcription of what the speaker actually uttered. Please note that the total number of chars in each row must not exceed 75 (80 including the characters needed for the mnemonic and the useful next blank).

“BSG:” begin segment specifies the starting point of segmented speech sequence. It is set to zero.

“ESG:” begin segment specifies the ending point of segmented speech sequence. Field present only in hand labelled speech files. Empty if file was not segmented.

“LBR:” row stores information about the acquisition window (LaBelling during Recording) and the prompt text, i.e. what the speaker should have uttered. The related fields are: sequence beginning (in sample), sequence end, input gain on recording, minimum sample value, maximum sample value, orthographic text prompt. Input gain, minimum and maximum values are left blank.

“LBO:” specifies a broad segmentation (LaBelling Orthographic) with the transcription of what the speaker actually said. The related fields are: sequence beginning (in sample), sequence centre, sequence end, corrected orthographic text. The sequence centre is left blank. Begin and end coincide with the “LBR:” end points.

“EXT:” text extension field allows orthographic texts to be continued on multiple rows

Blanks should not be treated as significant in label fields, except to separate words in the orthographic transcriptions. The same principle applies when merging “LBR:” and “EXT:” transcription text and a space is assumed to be inserted between these lines. However when performing processing of such lines, and later doing text comparisons, multiple blanks could arise, but white space should be considered insignificant. In any case, words may not be split over lines.

The orthographic texts are written using the ISO-8859-2 (Latin 2) character set.

mnem.	item format	example	comments
LBD:	“”	-	label body keyword
LBR:	“%lu, %lu, %d, %d, %d, %s”	0, 12457, , , , This is what	labelling during recording: begin, end, gain, min, max, orthographic text prompt
EXT:	“%.75s”	the speaker should have uttered	line extension

LBO:	"%lu, %lu, %lu, %s"	0, 6229, 12457, This is what the	orthographic labelling: begin, centre, end, orthographic transcription text
EXT:	"%.75s"	speaker actually said	line extension

If the speech signal is manual segmented, then the SAMPA transcription is present. Example for number "four" ("štiri") is presented:

BSG:

11	4304	951	si
4339	5383	134	S
5383	6715	103	t
6732	8526	301 503	i:
8526	9143	122	r
9143	11318	301	I
11330	14356	951	si

ESG:

First column presents consecutive sample number for beginning of phonem, next number presents ending of this phonem, then the IPA and the SAMPA codes for phoneme follow. The orthographic texts are written using the ISO-8859-2 (Latin 2) character set.

```
LHD: SAM, 5.10
DBN: SNABI_telephone
VOL: SI_SFN_01
SES:
CMT: *** Speech file information ***
DIR: /SFN/MMC/1B/spk_D3
SRC: D3MT1159.WSS
CCD: MMC
```

```

CRP:
BEG: 0
END: 57639
REP: UMARIB, MARIBOR, SLOVENIA
RED: 11/May/1995
RET:
CMT: *** Speech data coding ***
SAM: 16000
SNB: 2
SBF: MOTOROLA
SSB: 16
QNT: LIN
CMT: *** Speaker information ***
SCD: D3
SEX: M
AGE: 25
ACC: 3
CMT: *** Recording conditions ***
REG: 4
ENV: OFFICE
NET: PSTN
PHM: TOUCHTONE
LBD:
CMT: *** Transcription data ***
CMT: LBR: begin, end, input gain, min, max, prompt
CMT: LBO: begin, center, end, transcription
LBR: 0, 57639, , , , Izpiši neprebrana sporočila
LBO: 0, , 57639, Izpiši neprebrano sporočilo
LBP:
BSG:
ESG:

```

Figure 1 Example of a label file

2. Database design and collection

2.1 Recording site and platform

The database was recorded in a time span of 3 years. Majority of calls were made over the analogue telephone lines and several digital telephone lines. The speech was first recorded with DAT recorder and then transferred over the digital connection (using DAT-Link) to workstation, where it was stored with 16 kHz sampling rate and 16 bit linear quantization with MSB-LSB byte order.

2.2 Speaker recruitment

In contrast with the small vocabulary telephone speech databases that consists of hundreds or thousands of speakers, the selection of speakers in case of large vocabulary speech databases with several tenths of speakers can be more deliberate. In spite of that

we did not defined any special criteria according to which the speaker selection should be made. The goal in speaker selection was to achieve good balance according to gender and age categories. The majority of speakers were students and employees of the university. Speakers were from different dialect regions although we did not cover all the main dialect regions in Slovenia. During the recording the speakers were not instructed to use standard pronunciation but rather to use their usual pronunciation.

All recordings were done in a normal office acoustic environment avoiding exaggerated presence of sounds like door slam, background music or cross talk, paper rustle, etc.

The speakers read different groups of sub-corpora making short pauses after each sub-corpus. The isolated words, alphabet, and set of digit strings were considered as sub-corpora during the recording.

All together 82 speakers were recorded. From these 31 were female and 51 male speakers. Majority of the speakers were around 21 years old, the youngest was 15 years and the oldest 57 years old. Six out of ten main dialect regions in Slovenia were covered by speakers.

Each speaker uttered in average 200 sentences, 80 isolated words, containing also digits, 20 digit strings and alphabet. The database contains 22537 recordings of sentences, 5819 isolated words and 1504 recordings of numbers and number sets.

2.3 Design of prompting and prompt-sheet

The speakers read text from specially designed booklets with one item (word, letter, digit string or sentence) per page. In this way the speakers were able to concentrate only on the item uttered. They also did not try to hurry reading the text.

2.4 Transcription

The transcription used in this database is an orthographic lexical transcription. The transcription is intended to be a rough guide that users of the database can further examine for details.

All transcriptions were automatically checked for spelling errors and errors in transcription syntax.

For the transcriptions, the ISO-8859-2 character set was used. It is written in PDF file /SFN/DOC/ISO88592.PDF.

Transcriptions are case sensitive. Lexical items are written in small letters. Proper names, names of companies and geographical names start with capital letter accordingly to the Slovenian grammatical rules.

Lexical items have only one spelling and are spelled according to the Slovenian rules.

3. Database contents definition

The final specification for the Slovenian recordings is as follows:

Words:	
Abc	Alphabet (26 characters)
Numbers	Digits (12 digits)
Sets	20 digit strings
Wrđ	68 words (city names, command words)

MMC:	Divided in 3 sub-corpora, each sub-corpora was further divided in A and B sub-corpora; set of 529 sentences (from 5 – 25 words)
1A	96 sentences
1B	92 sentences
2A	84 sentences
2B	83 sentences
3A	86 sentences
3B	87 sentences
Lingua:	Divided in 4 sub-corpora, each sub-corpora was further divided in A and B sub-corpora; set of 911 sentences (from 3 – 25 words)
	112 sentences
1A	112 sentences
1B	112 sentences
2A	112 sentences
2B	112 sentences
3A	112 sentences
3B	118 sentences
4A	120 sentences
4B	Continuous speech - 90 contextual dependent sentences
Passage	

Table 4 - SNABI Slovenian database contents

In Table 5 the list of speakers and contents of recordings is presented. Means of presented mnemonics are:

Speaker – serial number and unique speaker code

Next three mnemonics present corpora and sub-corpora, which speaker pronounced. Details are explained in Table 3.

Words – include sub-corpora ABC, numbers, sets and words

MMC – include sub-corpora from 1A to 3B

Lingua – include sub-corpora from 1A to 4B and passage

ACC – speaker accent (Table 12, page 22),

REG – calling region (Table 13, page 22),

SEX – speaker gender,

AGE – speaker age (Table 14, page 23),

SMOK – presents if the speaker is smoker (S) or nonsmoker (N) (Table 16, page 23),

EDUC – degree of speaker education (Table 15, page 23),

speaker	words	MMC	lingua	ACC	REG	SEX	AGE	SMOK	EDUC
1 00	Wrd, numbers	1a, 1b		9	2	M	32	N	8
2 01	Wrd, numbers	1a, 1b		1	2	M	24	N	6
3 02	Wrd, numbers	1a, 1b		9	2	M	24	S	6
4 03	Wrd, numbers	1a, 1b		9	2	M	21	N	4
5 04	Wrd, numbers	1a, 1b		9	2	F	22	S	4
6 05	Wrd, numbers	1a, 1b		1	2	F	23	N	5
7 06	Wrd, numbers	1a, 1b		9	2	F	20	N	4
8 07	Wrd, numbers	1a, 1b		9	2	M	22	N	4
9 08	Wrd, numbers	1a, 1b		9	2	F	25	S	6

speaker		words	MMC	lingua	ACC	REG	SEX	AGE	SMOK	EDUC
10	09	Wrd, numbers	1a, 1b		9	2	F	19	N	4
11	0a	Wrd, numbers	1a, 1b		9	2	M	21	N	4
12	0b	Wrd, numbers	1a, 1b		9	2	M	22	N	4
13	0c	Wrd, numbers	1a, 1b		7	1	M	24	N	4
14	0d	Wrd, numbers	1a, 1b		2	3	F	22	S	4
15	0e	Wrd, numbers	1a, 1b		7	1	F	22	N	4
16	0f	Wrd, numbers	1a, 1b		3	4	F	23	N	4
17	0g	Wrd, numbers	1a, 1b		1	2	F	25	N	4
18	0h	Wrd, numbers	1a, 1b		3	4	F	22	N	4
19	0i	Wrd, numbers	1a, 1b		5	6	F	22	N	4
20	0j	Wrd, numbers	1a, 1b		1	2	M	20	S	4
21	0k	Wrd, numbers	1a, 1b		1	2	F	23	N	4
22	0l	Wrd, numbers	1a, 1b		1	2	M	21	N	4
23	0m	Wrd, numbers	1a, 1b		1	2	M	21	N	4
24	0n	Wrd, numbers	1a, 1b		1	2	M	22	N	4
25	0o	Wrd, numbers	1a, 1b		1	2	M	22	N	4
26	0p	Wrd, numbers	1a, 1b		1	2	M	25	N	5
27	0r	Wrd, numbers	1a, 1b		3	4	M	21	N	4
28	0s	Wrd, numbers	1a, 1b		3	4	M	21	N	4
29	0t	Wrd, numbers	1a, 1b		1	2	M	22	N	4
30	0u	Wrd, numbers	1a, 1b		1	2	F	29	N	8
31	0v	Wrd, numbers	1a, 1b		1	2	M	21	N	4
32	0z	Wrd, numbers	1a, 1b		1	2	M	21	N	4
33	10	Wrd, numbers	1a, 1b		1	2	M	57	N	8
34	11	Wrd, numbers	1a, 1b		9	2	M	47	N	8
35	12		1a, 1b		1	2	M	25	N	6
36	13	Wrd, numbers	1a, 1b		1	2	M	22	N	4
37	c1	Wrd, numbers, sets	2a	1b	5	6	F	20	N	5
38	c2	sets, abc	1a, 1b, 2b		1	2	M	19	N	4
39	c3	sets, abc	1a, 1b		7	1	M	20	N	4
40	c4	sets, abc	2a, 2b		2	3	M	22	N	4
41	c5	sets, abc	1a		5	6	F	21	N	4
42	c6	sets, abc	1a, 1b		6	7	F	19	N	4
43	c7	sets, abc	2a, 2b		6	7	F	20	N	4
44	c8	sets, abc	2a, 2b		5	6	M	25	S	4
45	c9		2a, 2b		1	2	F	22	N	4
46	ca	Wrd, numbers	3a, 3b		1	2	F	20	N	4
47	cb	Wrd, numbers	3a, 3b		6	7	F	25	N	6
48	cc	Wrd, numbers	3a, 3b		6	7	M	23	N	4
49	cd	Wrd, numbers	3a, 3b		1	2	F	21	N	5
50	ce	Wrd, numbers	3a, 3b		1	2	M	21	N	4
51	cf	Wrd, numbers	3a, 3b		2	3	M	20	N	4
52	cg	Wrd, numbers, sets, abc	2a, 2b		3	4	M	21	N	4
53	ch	Wrd, numbers, sets, abc	3a, 3b		1	2	F	23	N	4
54	ci	Wrd, numbers, sets, abc	3a, 3b		3	4	F	21	S	4
55	cj	Wrd, numbers, sets, abc	2a, 2b		2	3	M	21	N	4
56	ck	Wrd, numbers, sets, abc	2a, 2b		3	4	F	23	S	4
57	cl	Wrd, numbers, sets, abc	1a, 1b		7	1	F	24	S	4
58	cm	Wrd, numbers, sets, abc	1a, 1b		6	7	M	23	S	4

speaker		words	MMC	lingua	ACC	REG	SEX	AGE	SMOK	EDUC
59	co	Wrd, numbers, sets, abc		1a, 1b, 2a, 2b, passage	9	2	M	50	N	4
60	cp (d0)	Wrd, numbers, sets, abc	1a, 1b	1a, 1b, 4a, 4b, passage	1	2	M	28	N	6
61	cs	Wrd, numbers, sets, abc		3a, 3b, 4a, 4b	1	2	M	49	N	4
62	ct	Wrd, numbers, sets, abc	2a, 2b	2a, 2b, 3a, 3b, passage	1	2	F	26	N	6
63	cu (dj)	Wrd, numbers, abc, sets	3a, 3b	4a, 4b, passage	9	2	M	32	N	6
64	cv	Wrd, numbers, abc, sets	1a, 1b	1a, 1b, 2a, 2b, passage	9	2	F	26	N	5
65	cz	Wrd, numbers, abc, sets	2a, 2b	3a, 3b, passage	1	2	M	28	S	5
66	d0 (cp)	Wrd, numbers	1a, 1b		5	6	M	21	N	4
67	d1	Wrd, numbers	3a, 3b		7	1	M	21	N	6
68	d2	Wrd, numbers	3a, 3b		7	7	F	27	S	4
69	d3	Wrd, numbers	1a, 1b		3	4	M	25	N	4
70	d4 (di)	Wrd, numbers	3a, 3b		5	6	F	20	N	4
71	d5	Wrd, numbers, abc, sets	2a, 2b	1a, 1b, 4a, 4b, passage	1	2	M	25	N	6
72	d6	Wrd, numbers, sets, abc	1a, 1b	2a, 2b, 3a, 3b, passage	9	2	M	26	S	6
73	d7	Wrd, numbers, abc, sets,	1a, 1b	1a, 1b, passage	1	2	M	42	N	8
74	d8	Wrd, numbers, abc sets,	3a, 3b	1a, 1b, 4a, 4b, passage	1	2	M	24	N	6
75	d9	Wrd, numbers, abc, sets		1a, 1b, 4a, 4b, passage	1	2	M	23	S	6
76	da (d1)	Wrd, numbers, abc sets,	1a, 1b	3a, 3b, passage	9	2	F	15	N	2
77	db (d2)	Wrd, numbers, abc sets,	2a, 2b	3a, 3b, passage	9	2	M	23	N	6
78	dc	Wrd, numbers, abc, sets	2a, 2b	4a, 4b, passage	7	1	M	27	N	7
79	dd	Wrd, numbers, abc sets,	1a, 1b	1a, 1b, passage	9	2	M	32	N	7
80	de	Wrd, numbers, abc, sets	2a, 2b	2a, 2b, passage	7	1	F	28	N	6
81	df	Wrd, numbers, abc sets,	3a, 3b	3a, 3b, passage	1	2	M	51	N	4
82	di (d4)	Wrd, numbers, abc, sets	3a, 3b	2a, 2b, 3a, 3b, passage	1	2	M	24	N	4

Table 5 - SNABI Database contents

3.1 Alphabet

Slovenian language has no standard names for letters so speakers were not accustomed to spelling. Only letters from Slovenian language were used. The following table lists the used letters, names for the letters that appeared ordered by decreasing frequency of occurrence and number of occurrences.

Letter	Names	Number count
a	A	33
b	B BE BA BO EB	33
c	C CE CA CO EC	33
č	Č ČE	33
d	D DE DO DI	33
e	E	32
f	F EF FE	33
g	G GE GO EG GA	32
h	H HA HE HI HO EH	33
i	I	33
j	J JE EJ JA	33
k	K KA KE KO	33
l	L EL LE LI	33
m	M EM AM ME	33
n	N EN NE NO	33
o	O	33
p	P PE PA EP	32
r	R ER RA RE	33
s	S ES SA SE SO	33
š	Š EŠ ŠE ŠO	33
t	T TE ET TA TI	33
u	U UJ	33
v	V VE VA VO EV	33
z	Z ZE ZA ZO ZU	33
ž	Ž ŽE ŽA	33
x	IKS EKS	33

Table 6 - Frequency count for alphabet

3.2 Passage - phonetically rich sentences

For this item a set of 90 sentences was designed. The context used is computer technology and science in the future. Manual revisions were required to attain good coverage of less frequent phonemes. The following table shows frequency of occurrence of monophones.

Monophone	Frequency count
@	482
@:	0
@r	220
E	6701

E:	932
N	55
O	9889
O:	609
S	1474
W	701
Z	870
a	7130
a:	7565
b	1746
b_n	0
d	3492
d_l	52
d_n	270
e:	4405
f	523
g	1693
i	7447
i:	3799
j	5850
k	5299
l	4916
l'	18
m	3816
n	9208
n'	20
o	430
o:	2657
p	5322
r	6789
s	5227
t	6447
tS	2735
t_l	15
t_n	195
ts	1125
u	1614
u:	1438
v	4003
w	1398
x	1946
z	2479
	$\Sigma = 133002$

Table 7 - Frequency of occurrence of monophones in phonetically rich sentences

3.3 Words

Speakers read 68 words. Corpus Words contains city names and command words. Manual revisions were required to attain good coverage of less frequent phonemes.

Monophone	Frequency count
@	146
@:	73
@r	73
E	881
E:	149
N	0
O	950
O:	365
S	368
W	217
Z	0
a	2035
a:	1527
b	944
b_n	0
d	950
d_l	0
d_n	0
e:	1169
f	73
g	435
i	1456
i:	1166
j	1237
k	1453
l	1383
l'	0
m	804
n	1973
n'	0
o	218
o:	1019
p	1383
r	2039
s	1533
t	1245
tS	586
t_l	0
t_n	0
ts	72
u	291
u:	726
v	801
w	0
x	291

z	793
	$\Sigma = 30824$

Table 8 - Frequency of occurrence of monophones in words

3.4 Numbers and number sets

Speakers read numbers from 0 to 9. Two numbers (“0” and “2”) in Slovenian language can be pronounced in two versions. Each speaker pronounced 11 numbers. In corpora Sets speaker pronounced 20 digit strings of length 6 digits.

Monophone	Frequency count
@	0
@:	0
@r	0
E	1456
E:	896
N	0
O	395
O:	0
S	695
W	0
Z	0
a	71
a:	729
b	0
b_n	0
d	1467
d_l	0
d_n	0
e:	1753
f	0
g	0
i	501
i:	1198
j	0
k	0
l	71
l'	0
m	827
n	1230
n'	0
o	0
o:	0
p	530
r	831
s	1155
t	2054
tS	501

t_l	0
t_n	0
ts	0
u	0
u:	71
v	1035
w	0
x	0
z	0
	$\Sigma = 17466$

Table 9 - Frequency of occurrence of monophones in numbers and number sets

3.5 MMC corpus

For this item a set of 529 sentences was designed. Sentences cover man-machine communication area. One part of sentences is taken from the field of train table information queries. Manual revisions were required to attain good coverage of less frequent phonemes. The following table shows frequency of occurrence of monophones.

Monophone	Frequency count
@	2687
@:	44
@r	1092
E	24044
E:	3768
N	15
O	30618
O:	3656
S	5440
W	2115
Z	2387
a	20068
a:	24554
b	8241
b_n	0
d	13747
d_l	0
d_n	946
e:	16890
f	1295
g	4123
i	21602
i:	21332
j	13799
k	23548
l	19381
l'	0
m	14261

n	25226
n'	57
o	1664
o:	10026
p	20045
r	25906
s	19241
t	21762
tS	6013
t_l	92
t_n	950
ts	3719
u	3405
u:	4236
v	11985
w	8181
x	2437
z	15346
	$\Sigma = 459944$

Table 10 - Frequency of occurrence of monophones in MMC corpus

3.6 Lingua

For this item a set of 911 sentences was designed. Sentences are taken from different styles of text, such as books and newspapers. The following table shows frequency of occurrence of monophones.

Monophone	Frequency count
@	1225
@:	21
@r	555
E	12711
E:	2157
N	107
O	17181
O:	1583
S	3082
W	1103
Z	1637
a	11287
a:	14307
b	4858
b_n	21
d	7883
d_l	168
d_n	595
e:	12942

f	422
g	3597
i	13935
i:	9362
j	12441
k	9490
l	10944
l'	166
m	9117
n	15987
n'	106
o	777
o:	5608
p	9591
r	12786
s	12752
t	12027
tS	4293
t_l	110
t_n	439
ts	1778
u	2223
u:	2410
v	7833
w	4049
x	3141
z	5579
	$\Sigma = 264388$

Table 11 - Frequency of occurrence of monophones in Lingua

4. Speaker demographic information

4.1 Accent/Regions

Slovenia is a republic with approximately 2 million inhabitants. According to data from bibliography (Geodetski zavod Slovenije: 'Karta slovenskih narečij'1993), Slovenia is divided in 10 main accent regions.

Number	Name of accent	Label in database	Number of speakers	Number of speakers (%)
1	Štajersko	1	33	40,24
2	Koroško	2	4	4,88
3	Gorenjsko	3	8	9,76
4	Notranjsko	4	0	0
5	Dolenjsko	5	6	7,32
6	Primorsko	6	5	6,10
7	Prekmursko	7	8	9,76
8	Rovtarsko	8	0	0
9	Maribor	9	18	21,95

10	Ljubljana	0	0	0
----	-----------	---	---	---

Table 12 - Distribution of speakers over accent

Number	Name of region	Label in database	Number of speakers	Number of speakers (%)
1	Pomurje	1	8	9,76
2	Štajersko	2	51	62,20
3	Koroško	3	4	4,88
4	Gorenjsko	4	8	9,76
5	Notranjsko	5	0	0
6	Dolenjsko	6	6	7,32
7	Primorsko	7	5	6,10

Table 13 - Distribution of speakers over regions

4.2 Speaker characteristics

Age groups	Label in database	Number of male speakers	Number of female speakers	Percentage of total
under 16		0	1	1,22
16-30		42	30	87,80
31-45		4	0	4,88
46-60		5	0	6,10
over 60		0	0	0

Table 14 - Distribution of speakers over age groups and genders

Educational level	Label in database	Number of speakers	Percentage of total
Elementary school	2	1	1,22
Vocational school	3	0	0
Secondary school	4	53	64,63
College	5	6	7,32
University	6	15	18,29
Master's degree	7	2	2,44
Doctor's degree	8	5	6,10

Table 15 - Distribution of speakers over educational level

Smoking habits	Label in database	Number of speakers	Percentage of total
Nonsmoker	NONSMOKER	68	82,93
Smoker	SMOKER	14	17,07

Table 16 - Distribution of speakers over smoking habits

The file .../DOC/SPEAKER.TBL contains information about the accent region, gender, age, educational level and smoking habits of speakers in the database.

5. The lexicon

The lexicon is included in file /SFN/TABLE/LEXICON.TBL.

The lexicon file is an alphabetically ordered list of distinct lexical items (essentially words in our case) which occur in the corpus with the corresponding pronunciation information. Each distinct word has a separate entry. As the lexicon is derived from the corpus it uses the same alphabetic encoding for special and accented characters as used in the transcriptions (ISO-8859-2).

We have considered it useful to include a frequency count for each entry in the lexicon e.g. to help indicate rare words whose transcriptions are perhaps less important or reliable.

As the transcription activity is not phonetically based, we cannot accurately identify occurrences of homographs or heterophones, and so cannot provide frequency counts for these items; in fact the frequency count is orthographically based and not phonemically based.

Phonetic transcriptions in the lexicon were produced automatically with a grapheme-to-phoneme converter developed at the University of Maribor. Transcriptions were manually checked. Transcriptions use SAMPA phonetic alphabet. List of used symbols is stored in file /SFN/DOC/SAMPALEX.PDF.

Entries are case sensitive and are split only at spaces.

The following allophonic variations can be merged into one symbol:

Monophone	Allophonic variations
p	p p_n
b	b b_n
t	t t_n t_l
d	d d_n d_l
n	n N
l	l
e	e e:
E	E E:
a	a a:
O	O O:
i	i i:
u	u u:
o	o o:
@	@ @r
w	w W
tS	tS dZ
ts	ts
x	x

Table 17 - Allophonic variations of phonemes

6. Recording conditions

6.1 Environments

Environment	Label in database	Number of calls	Percentage
Home	HOME	0	0
Office	OFFICE	82	100
Public place	PUBLIC	0	0
Telephone booth	BOOTH	0	0
Factory	FACTORY	0	0
Other	OTHER	0	0

Table 18 - Number of calls and percentage as a function of the calling environment.

6.2 Network

All calls were made from the public fixed telephone network.

6.3 Handsets

Handset	Label in database	Number of calls	Percentage
Set with touchtone dialing	TOUCHTONE	82	100
Set with rotary dialing	ROTARY	0	0
Cordless set	CORDLESS	0	0
Other	OTHER	0	0

Table 19 - Number of calls and percentage as a function of the handset used.

7. Test set definition

The primary task within the SNABI project is the provision of spoken language resources recorded over telephone lines and in studio environment. It is recognised that subsequent application of these resources will involve recogniser development and evaluation which would benefit from the existence of standardised evaluation procedures including the definition of test sets for comparative evaluation.

It has been decided to meet this requirement by supplying each database with a relatively simple test set definition enabling comparative evaluation within the most obvious tasks for which the databases are intended. These test sets are only meant to be recommendations for meaningful tests to be conducted, and users are of course free to

use the data for their own purposes or to propose, to the community, additional test sets for further standardised testing.

7.1 Selection of test speakers

The selection of test speakers takes into account that the remaining speakers are likely to be applied for training and it has to aim at providing a material that will result in statistically significant results.

N speakers constituting the test set were selected using a controlled random selection procedure in which cells were created for each combination of region and gender. The number of speakers belonging to each cell was then reduced by a factor of N/F, F being the total number of speakers in the database. For the second test set the text was selected first and then the speakers were selected.

The reduction within each cell was done by random exclusion of speakers until the required integer number has been reached.

N is 7 for Slovenian Fixed Network Database.

The statistical significance varies with the number of items within each test.

7.2 Test Set Specifications

The following sections detail the requirements to the test set specifications for SNABI database.

The selection of test speakers is divided in two groups. The first group is selected in a way, that test set is speaker independent. Second group is not only speaker independent, but also textual independent. This training set can be used for adaptation of language models.

The sessions used for testing were selected according to a controlled random selection procedure (see previous section).

7.2.1 Test session list file (*.SES)

File: /SFN/INDEX/F1TSTSI.SES

The test session list file serves as an overall specification of the sessions/speakers actually used for testing. The following filename convention is used for the test session list file:

<DD> TST <LL> . SES

where:

<DD>	Database identification code (00-ZZ) For SNABI: F1, F2=fixed net, S1, S2=studio
<LL>	Modified ISO 639 two-letter language code (e.g. SI=Slovenian)

Table 19 - Test session list file name convention

Each line in this file contains a session code spk_NN, as shown in the example below.

spk_DA

```

spk_C1
spk_CU
spk_C9
spk_0F
spk_0D
. . . .

```

Figure 2 - Example of a test session list file

A number of specific tests designs should be specified for every database. As a minimum all corpus codes must be covered in at least one test. Every test is described in a test description file:

7.2.2 Test description file (*.DSC)

Files: /SFN/TEST/SI<XX>SI.DSC

The test description file gives the information needed to describe a particular test. The following filename convention is used for test description files

<DD> XX <LL> . DSC

where:

<DD>	Database identification code (00-ZZ) For SNABI: F1, F2=fixed net, S1, S2=studio
XX	Unique test identifier
<LL>	Modified ISO 639 two-letter language code (e.g. SI=Slovenian)

Table 20 -Test description file name convention

The test description file is a SAM file where the following mnemonics are mandatory

TST: <XX> (Test Identifier)

Specifies start of a test description file. XX denotes a zero-extended two-digit code.

CCD: <C1> [<C2>][<C3>]... [<CN>]

Specifies corpus codes involved.

TSF: DDSXXLL.LST (Test list file)

Files: /SFN/TEST/A1S<XX>SI.LST

The test list file contains a list of full path names of speech label files to be used in test XX.

A listing file contains all label files from the defined test sessions and the selected corpus. Label files are listed, one per line, including full path information, in the following format:

VOLUME / DATABASE / CORPUSN / SPEAKERID / DDNNNNCC . WSS

TSZ: n (Test set size)

Number of lines in the test list file <DD>S<XX><LL>.LST.

VCF: DDVXXLL.VCB (Test vocabulary file name)

Files: /SFN/TEST/A1V<XX>SI.VCB

The test vocabulary is listed in the file <DD>V<XX><LL>.VCB, with one line for each vocabulary entry. Vocabulary entries can be lexicon words or phrases (lexicon words separated by single spaces). The file is stored in the INDEX directory.

It should be noted that the test vocabulary is not necessarily the same vocabulary used by a typical recogniser under test. For "open" vocabulary tasks, such as names and large vocabulary continuous speech recognition, the recogniser's vocabulary will generally be larger than the one tested, but still not containing all testset words. However, for "closed" vocabularies, such as digits, numbers and application words, the test vocabulary is typically also chosen for the recogniser.

A frequency count is included for every vocabulary entry. This will show the testset coverage and facilitate confidence estimation for individual word recognition rates. The count should equal the number of occurrences of each entry in the preprocessed testset reference transcriptions.

The frequency count is separated from the vocabulary entry by a TAB character.

SCM: [SENTENCE MATCH | WORD MATCH] (scoring method)

This field specifies the scoring method that must be applied with the test. Different scoring methodologies are explained below.

The task of a scoring procedure is to compare the recogniser output for all test set utterances to the correct sentence transcriptions, as defined in the label files. To achieve comparable results between different test sets, it is essential that the exact scoring method for each test is described as part of the test design.

Scoring is conveniently implemented in two stages. The first stage is a pre-processing where the reference and test transcriptions are converted to strings of scoring symbols. The second stage is the actual comparison of test and reference strings, including error classification and counting.

The pre-processing of the test transcription (recogniser output) is not a part of the test design, since it is specific to the recogniser. However, the recogniser output should be matched to the scoring algorithm so that the number of recognition errors are minimised.

The pre-processing of reference transcriptions is an essential part of the test design, and needs to be documented accurately, along with the error counting method.

Two standard scoring methods have been defined for the SNABI database tests; a sentence match intended for isolated word or phrase recognition, and a word match based on dynamic programming (DP) alignment. The second method can be used both for

continuous speech recognition and keyword spotting tests. Any deviations from these two default scoring methods should be explicitly described in the test design.

Sentence Match

Errors are counted by comparing the pre-processed test and reference transcriptions, and counting the following four categories:

- C: Correctly recognised utterances. Test and reference strings are exactly equal. This category includes correct rejections, where both are empty.
- S: Substitution errors. Both test and reference are vocabulary entries, but not the same.
- FR: False rejection errors. The reference is a vocabulary entry, but the recognition output is empty.
- FA: False acceptance errors. The reference is empty, but a vocabulary entry has been recognised.

Word Match

The NIST dynamic programming algorithm should be used to align the pre-processed reference and test strings. Results are presented as the following five numbers:

- D: Deletion errors
- S: Substitution errors
- I: Insertion errors
- C: Correctly recognised words
- CS: Correctly recognised sentences

VSZ: n (Test vocabulary size)

The number of vocabulary entries equals the number of lines in the test vocabulary file <DD>V<XX><LL>.VCB.

ACN: n (Number of acceptable utterances)

This is the total number of vocabulary entries in the preprocessed reference transcriptions. It is equal to the sum of frequency counts in the vocabulary file. The number can be used to compute error and recognition rates, as well as test confidence intervals.

RCN: n (Number of rejectable utterances)

With sentence match scoring, this is the number of empty reference utterances after the preprocessing. $RCN + CAN = TSZ$. For word match scoring, the rejection count is the number of symbols deleted by the preprocessing step, i.e. word fragments.

ETS:

Specifies the end of a test description file.

7.3 Tests defined for SNABI SFN database

The following tests were defined for SNABI Slovenian Fixed Network database:

Session	Test identifier	Corpus items	Word/sentence match	Vocabulary size
F1	01	MMC	word	1022
F1	02	lingua	word	3144
F1	03	words/abc	sentence	26
F1	04	words/sets	word	10
F1	05	words/numbers	sentence	12
F1	06	words/wrd	sentence	68
F2	01	lingua	word	1133

Table 21 -Tests defined for SNABI SFN database

Appendix A

8. Set of used sentences for corpus MMC

8.1 Corpus MMC 1A

Shrani datoteko kot besedilo (M1002)
Preidi v brošuro (M1003)
Koliko pomnilnika je še na razpolago (M1004)
Iz označenih podatkov tvori tabelo (M1005)
Katera je klicna številka za Francijo (M1006)
Prikaži seznam posebnih znakov (M1007)
Nujno me zvežite s policijo (M1008)
Zvežite me, prosim, z bolnišnico (M1009)
Prikaži sintakso ukaza za kopiranje datotek (M1010)
Skoči na začetek strani (M1011)
Barvo ozadja spremeni v črno (M1012)
Iztiskaj vsebino okna dve v datoteko iztisk pika P R N (M1013)
Ali je gonilnik za miško naložen (M1014)
Zapusti urejevalnik (M1015)
Želim, da vse, ki me pokličejo od petindvajsetega maja dalje, obvestite o moji odsotnosti (M1016)
Vklopi radio (M1017)
Pripravi uporabo kodnih tabel (M1018)
Izpiši brošure (M1019)
Pokaži sistemski čas (M1020)
Izpiši sporočila, ki so prejeta po datumu (M1021)
Zatemni retrovizorsko ogledalo (M1022)
Priredi tipki ukaz in ga izvedi takoj po pritisku nanjo (M1023)
Postavi tekoče sporočilo v urejevalnik besedila (M1024)
Shrani datoteko pod imenom tekst pika doc (M1025)
Postavi psevdonim (M1026)
Prikaži vse tipe znakov (M1027)
Skoči v vrsto sedeminosemdeset (M1028)
Kakšna je obremenitev sistema (M1029)
Ustvari datoteko za shranjevanje nastavitev tipk (M1030)
Posreduj sporočilo dalje (M1031)
Kakšno je stanje tiskalnika (M1032)
Vključi besedilo urejevalnika Word Star (M1033)
Preidi na naslednje sporočilo (M1034)
Tvori imenik (M1035)
Zbriši datoteko (M1036)
Prišlo je do razlike v številu porabljenih impulzov med mojim in vašim števcem (M1037)
Zbriši zaslon (M1038)
Želim klicati v Kanado na račun klicanega (M1039)
Izpiši seznam procesov (M1040)
Želim bujenje ob peti uri in dvajset minut na telefonski številki ena nič nič dve dve ena (M1041)
Premakni vzglavnik voznikovega sedeža za tri centimetre navzgor (M1042)
Odklopite naročnika s telefonsko številko dve nič dve sedem sedem ena do petnajstega avgusta tisoč devetsto triindevetdeset (M1043)
Je možno vzpostaviti zvezo med tremi telefonskimi naročniki hkrati (M1044)
Vstavi tri tabulatorje (M1045)
Želim, da zamenjate telefonski aparat (M1046)
Ali na Prešernovi ulici stanuje kakšen Krajnc in na kateri hišni številki (M1047)
Formatiraj disk (M1048)
Spremeni dimenzije lista (M1049)
Sodim stranem priredi komentarje (M1050)
Uredi besedilo v dva stolpca (M1051)
Prekini elektronsko pošto (M1052)
Premakni krmilo nekoliko naprej (M1053)
Izpiši drevesno strukturo seznamov (M1054)
Kako zapustim program (M1055)

Seznam vseh sporočil vnesi v urejevalnik besedila (M1056)
 Izpiši seznam vseh sporočil (M1057)
 Me lahko seznaniš z odhodi vlakov proti Ljubljani v popoldanskem času (M1058)
 Shrani sporočilo v datoteko - brez glave (M1059)
 Spremeni način delovanja tiskalnika (M1060)
 Preklopi na grafični način (M1061)
 Zapusti elektronsko pošto (M1062)
 Izpiši vse podatke o sporočilu (M1063)
 Vključi prikaz koordinat položajev kazalca in miške (M1064)
 Označi sporočilo (M1065)
 Na vogalu Prešernove in Maistrove ulice se je zgodila prometna nesreča (M1066)
 Odgovori in vrni kopijo sporočila (M1067)
 Prestavi sporočilo v brošuro (M1068)
 Vključi objekt iz AutoCada (M1069)
 Pokaži datum in čas (M1070)
 Navzdol poišči niz (M1071)
 Iztiskaj datoteko (M1072)
 Zbriši zaslon (M1073)
 Vključi grelec zadnjega okna (M1074)
 Odpri zadnje levo okno (M1075)
 Izpiši sporočila, ki so prejeta po datumu (M1076)
 Ustvari rezidenčni proces in počakaj na izvršitev (M1077)
 Ponovi prejšnji ukaz (M1078)
 Odpri novo okno (M1079)
 Zapri streho (M1080)
 Pokaži primer izvajanja ukaza za brisanje zaslona (M1081)
 Označi besedilo in spremeni tip znakov (M1082)
 Povejte mi rezultat nogometne tekme med Mariborom in Koprom (M1083)
 Koliko stane priključitev internega števca impulzov (M1084)
 Ko dvignem slušalko, slišim pogovore drugih (M1085)
 Ali lahko, prosim, poveste, kdaj pripelje danes zadnji avtobus iz Kopra (M1086)
 Povečaj prikaz na zaslonu (M1087)
 Nizu priredi ukaz (M1088)
 Izberi besedilo do konca zapisa (M1089)
 Skrči tekoče okno na minimum (M1090)
 Format strani naj bo A štiri (M1091)
 Preveri slovnično pravilnost in popravi napake (M1092)
 V skladišču je prišlo do požara (M1093)
 Prijavljam telefonsko okvaro (M1094)
 Abecedni seznam ukazov (M1095)
 Potrebujem pomoč (M1096)
 Kopiraj disk (M1097)

8.2 Corpus MMC 1B

Zakleni vrata (M1098)
 Poženi motor (M1099)
 Prikaži začetek tekočega poročila (M1100)
 Prikaži prvo sporočilo (M1101)
 Kdo je avtor tega programa (M1102)
 Preimenuj datoteko (M1103)
 Vzpostavite nujen telefonski pogovor s koprsko obalno stražo (M1104)
 Odkleni prtljažnik (M1105)
 Vkljopi pozicijske luči (M1106)
 Izvleci telefonsko anteno (M1107)
 Podčrtaj tekočo vrstico (M1108)
 Mi lahko pojasnite potek vzpostavljanja telefonskih zvez z Madagaskarjem (M1109)
 Skoči na konec besedila (M1110)
 Spremeni dolžino tabulatorjev (M1111)
 Levi rob pomakni za šest centimetrov na desno (M1112)
 Izpiši celotno stran (M1113)
 Odgovori na zadnje sporočilo (M1114)

Želim vložiti prošnjo za spremembo telefonske številke (M1115)
 Piši poudarjeno (M1116)
 Tvorí sklop ukazov (M1117)
 Previj kaseto (M1118)
 Razširi tekoče okno na celoten zaslon (M1119)
 Prestavi datoteko (M1120)
 Zbij datoteko (M1121)
 Vstavi tretji C D (M1122)
 Kako naj poženem razhroščevalec (M1123)
 Povežite me s telefonsko številko enaindvajset osemsto pet v Ljubljani (M1124)
 Odgovori z vsebino mojega zadnjega sporočila (M1125)
 Združi besedilo z vsebino datoteke združitev pika T X T (M1126)
 Ali lahko naročim medkrajevni telefonski pogovor (M1127)
 Odgovori z osebnim imenom (M1128)
 Želim, da mi danes odklopite telefon, ker me venomer kliče neki neznanec (M1129)
 Bi lahko izvedel, s katerih telefonskih števil je bila v preteklih petih urah vzpostavljena telefonska zveza (M1130)
 Zadnje telefaks sporočilo vključi v datoteko (M1131)
 Izpiši vsebino seznama (M1132)
 Naslove piši poudarjeno z velikostjo petnajst milimetrov (M1133)
 Že dva dni ne morem vzpostaviti zveze z Ribniško kočó na Pohorju (M1134)
 Nastavi frekvenco petindevetdeset cela devet megahertzov (M1135)
 Logično poveži enoto s seznamom (M1136)
 Poravnaj z levim robom (M1137)
 Kopiraj datoteko (M1138)
 Odgovori z vsebino sporočila, na katero odgovarjam (M1139)
 Preveri, ali je tiskalnik priključen (M1140)
 Obnovi zbrisano datoteko (M1141)
 Kakšen je pomen napake pod zaporedno številko sedemnajst (M1142)
 Me lahko seznaniš s snežnimi razmerami, temperaturo in obratovanjem vlečnic na mariborskem Pohorju (M1143)
 Povejte mi, prosim, telefonsko številko stanovanja Cankar na Vrhnikih (M1144)
 Kopiraj skupino datotek (M1145)
 Izpisi datoteko na L P T ena (M1146)
 Poišči sporočilo, ki vsebuje niz (M1147)
 Prestavi sporočilo v brošuro datoteke (M1148)
 Ali mi lahko poveste, čigava je telefonska številka dve dve tri šest pet devet v Mariboru (M1149)
 Pomakni sovoznikov sedež nekoliko nazaj (M1150)
 Opozorjam na nevarnost naravne nesreče (M1151)
 Izklopi dolge luči in meglenke (M1152)
 Preveri stanje diska (M1153)
 Razmik med vrsticami postavi na pet cela tri centimetra (M1154)
 Spremeni znak pripravljenosti (M1155)
 Spremeni barvo znakov v zeleno (M1156)
 Primerjaj vsebini datotek (M1157)
 Izključi alarm (M1158)
 Izpiši neprebrana sporočila (M1159)
 Lahko pošljem telegram preko telefona (M1160)
 Pokliči telefonsko številko dve ena ena osem osem nič (M1161)
 Izpiši uporabnike mreže (M1162)
 Želim izvedeti ceno in trajanje impulza za klic v Avstrijo (M1163)
 Zapri prednji okni (M1164)
 Obnovi formatiran disk (M1165)
 Uredi zadnje sporočilo (M1166)
 Oštevilči lihe strani (M1167)
 Preberi sporočilo (M1168)
 Skopiraj sporočilo v brošuro (M1169)
 Poišči datoteko (M1170)
 Spremeni seznam (M1171)
 Izpiši označena sporočila (M1172)
 Ustvari rezidenčni proces (M1173)
 Zbriši vsa sporočila (M1174)

Izpiši sporočila, ki vsebujejo niz (M1175)
 Pošlji sporočilo (M1176)
 Zbriši podseznam (M1177)
 Vključi vse smerokaze (M1178)
 Vključi brisalce na položaj dve (M1179)
 Koliko stane pet minut razgovora z Adis Abebo v Etiopiji (M1180)
 Dajte mi, prosim, telefonsko številko gasilcev na Teznu (M1181)
 Vzpostavi stanje pred izvedbo zadnjega ukaza (M1182)
 Iztiskaj sporočilo (M1183)
 Obnovi poškodovane datoteke (M1184)
 Določi tipko za izvedbo ukaza (M1185)
 Izpiši besedilo v eni barvi (M1186)
 Postavi značko na začetek datoteke (M1187)
 Malo črko spremeni v veliko črko (M1188)
 Segrej vžigalnik (M1189)

8.3 Corpus MMC 2A

Pozdravljeni, bo čez dva dni peljal spalni vlak iz Ljubljane v Pariz (M1190)
 Je še kaka hitrejša možnost (M1191)
 Ali lahko dobim kak popust (M1192)
 Kolikšna pa je cena karte za potniški vlak iz Ormoža do Čakovca (M1193)
 Kolikšna je cena sezonske vozovnice na relaciji Pragersko-Celje (M1194)
 Dober dan, rad bi vedel, ob kateri uri bo šel direktni vlak za Muenchen (M1195)
 Dobro jutro, prosim, ali gre kak direktni vlak iz Pragerskega do postaje na Dunaju (M1196)
 Dober dan, povejte mi prosim, ali bo odpeljal vlak proti Krškemu (M1197)
 Pozdravljeni, zanima me, ali je kak vlak iz Trbovelj za Kranj (M1198)
 Mi lahko poveste, prosim, koliko zamude je imel potniški vlak iz čakovca (M1199)
 Bo vlak iz Novega mesta zamudil (M1200)
 Je imel poslovni vlak iz Sevnice zamudo (M1201)
 Rad bi zvedel, ali je Savinja ekspres z Jesenic prišel z zamudo (M1202)
 Dober dan, ima brzovlak iz Celja zamudo (M1203)
 Katere vrste vlak je tisti, na katerega bom prestopil (M1204)
 Dober večer, zanima me, ali obstaja direktna zveza za Amsterdam (M1205)
 Kje je treba prestopiti (M1206)
 Želim sedeti ob oknu (M1207)
 Ima vlak restavracijo (M1208)
 Hvala (M1209)
 Za kadilce (M1210)
 Kdaj pelje danes vlak s postaje v Sevnici proti Postojni (M1211)
 Je to najcenejša relacija (M1212)
 Je možen popust (M1213)
 Dober dan, rad bi vedel kolikšna je cena povratne karte za prvi razred za vlak Pohorje ekspres iz Laškega do Trbovlja (M1214)
 Koliko stane enosmerna karta za ekspresni vlak do Postojne (M1215)
 Dober dan, ali gre danes zvečer kak vlak za Budimpešto (M1216)
 Dober večer, zanima me, ali je odpeljal vlak za Zagorje (M1217)
 Ali je že odpeljal vlak s postaje v Ormožu proti Lendavi (M1218)
 Ob kateri uri gre vlak s postaje Pragersko v smeri Dobova (M1219)
 Dober dan, koliko zamude ima vlak iz Bruslja (M1220)
 Zanima me, ali ekspresni vlak iz Ormoža zamuja (M1221)
 Ima kaj zamude (M1222)
 Dober večer, bo IC iz Kranja pripeljal z zamudo (M1223)
 Je imel potniški vlak iz Krškega kaj zamude (M1224)
 Katere vrste je vlak, na katerega moram presesti (M1225)
 Do kdaj je potrebno čakati na zvezo (M1226)
 Bom moral na relaciji Lendava-Budimpešta tudi prestopiti (M1227)
 Lahko plačam s kreditno kartico (M1228)
 Mi lahko pošljete predračun za vozovnico Maribor - Koper (M1229)
 Železniške informacije (M1230)
 Ob kateri uri odpelje (M1231)
 Koliko stane vozovnica (M1232)

Ne (M1233)
 Kolikšna je cena skupinske vozovnice (M1234)
 Kolikšna je cena mesečne vozovnice za relacijo Celje - Maribor (M1235)
 Lahko izvem ali je na vlaku za Postojno potrebno kje prestopiti (M1236)
 Kdaj imam prvo zvezo iz Maribora za Berlin (M1237)
 Kolikokrat bom moral prestopati (M1238)
 Ni kakšne zveze z manj čakanja med prestopom (M1239)
 Bo imel vlak iz Zidanega Mosta zamudo (M1240)
 Ima Zeleni vlak iz Kidričevega zamudo (M1241)
 Je bilo kaj zamude (M1242)
 Dober dan, lahko izvem, koliko zamude je imel vlak iz Pariza (M1243)
 Pozdravljeni, rad bi vedel, ob kateri uri odpelje Intercity vlak za Koper (M1244)
 Prosim povejte mi, ali je poslovni vlak proti Dunaju že odpeljal (M1245)
 Dobro jutro, bo šel poslovni vlak iz Maribora do Jesenic (M1246)
 Dober večer, zanima me, ali pelje poslovni vlak za Trst (M1247)
 Dober dan, koliko stane enosmerna karta za brzovlak iz Maribora do Kidričevega (M1248)
 Koliko stane vozovnica za prvi razred brzovlaka iz Lendave do Murske Sobote (M1249)
 Lahko dobim kak popust za otroke (M1250)
 Je potrebno še kaj doplačati (M1251)
 Dobro jutro, zanima me, ob kateri uri bo šel danes brzovlak za Trbovlje (M1252)
 Kolikšna je cena vozovnice (M1253)
 Ja (M1254)
 Kdaj prispe vlak iz Celja (M1255)
 Želim povratno vozovnico (M1256)
 Zanima me, ali moram z vlaka za Gradec prestopiti (M1257)
 Rad bi vedel, kako je z zvezami za Divačo (M1258)
 Bom moral veliko presedati (M1259)
 Ne želim prestopati (M1260)
 Dober večer, zanima me, ali bo Pohorje ekspres prišel z zamudo (M1261)
 Dober večer, bo imel vlak iz Gradca zamudo (M1262)
 Je kaj zamude (M1263)
 Povejte mi prosim, koliko zamude ima potniški vlak iz Lendave (M1264)
 Dober dan, mi lahko poveste kdaj je odpeljal kak vlak proti Kranju (M1265)
 Pozdravljeni, ali bo šel direktni vlak iz Kopra proti Dunaju (M1266)
 Dober dan, ob kateri uri pelje zeleni vlak iz Maribora proti Celju (M1267)
 Rad bi vedel, kdaj pelje kakšen vlak proti Ljubljani (M1268)
 Dobro jutro, želim zvedeti kolikšna je cena povratne vozovnice za zeleni vlak iz Zidanega Mosta do Kopra (M1269)
 Koliko stane tedenska vozna karta do Ormoža (M1270)
 Lahko dobim popust za upokoјence (M1271)
 Moram še kaj doplačati (M1272)
 Prosim, povejte mi, ali bo šel danes ob petih spalni vlak za Amsterdam (M1273)

8.4 Corpus MMC 2B

Povejte mi, prosim, ob kateri uri pelje vlak iz Maribora za Ljubljano (M1274)
 Kje vse moram prestopiti (M1275)
 Kdaj pelje vlak v popoldanskem času (M1276)
 Koliko časa moram čakati v Zidanem Mostu (M1277)
 Ne obstaja kakšna boljša zveza (M1278)
 Na kateri postaji moram presesti na drugi vlak (M1279)
 Oprostite, kako dolgo moram čakati (M1280)
 Kako bo med presedanjem z mojo prtljago (M1281)
 Lahko izvem, ali bo imel ekspresni vlak iz Postojne kaj zamude (M1282)
 Je Emona ekspres iz Trbovelj prispel z zamudo (M1283)
 Bo imel kaj zamude (M1284)
 Dober večer, rad bi vedel, ali je IC vlak iz štor pozen (M1285)
 Koliko zamude ima vlak iz Ormoža (M1286)
 Zdravo, zanima me, ali je že odpeljal potniški vlak za Zagorje (M1287)
 Ob kateri uri pelje kak vlak iz smeri Zidani Most proti Celju (M1288)
 Dober dan, ob kateri uri pelje brzovlak v Koper (M1289)
 Pozdravljeni, lahko izvem, kdaj gre jutri zvečer brzovlak za Kranj (M1290)

Koliko stane najcenejša varianta potovanja z Jesenic v Prago (M1291)
 Kakšna je cena tedenske vozovnice za prvi razred poslovnega vlaka na relaciji Pragersko-Hrastnik (M1292)
 Obstaja popust za družinska potovanja (M1293)
 Dober večer, ali pelje jutri zjutraj ekspresni vlak za Trst (M1294)
 Je to ekspresni vlak (M1295)
 Da (M1296)
 Je mogoče dobiti popust (M1297)
 Kdaj vozijo brzovlaki iz Ljubljane za Koper (M1298)
 Kakšne so možnosti potovanja z vlakom v Muenchen (M1299)
 V katerem kraju je treba prestopiti (M1300)
 Kako dolgo bom moral počakati na zvezo (M1301)
 Na kateri peron pripelje drugi vlak (M1302)
 Pozdravljeni, povejte mi, prosim, ali je imel poslovni vlak iz Maribora kaj zamude (M1303)
 Dober dan, ali je imel vlak iz smeri Dunaj ob prihodu kaj zamude (M1304)
 Je imel kaj zamude (M1305)
 Ali vlak iz Budimpešte zamuja (M1306)
 Dober dan, povejte mi, prosim, ob kateri uri pelje vlak za Maribor (M1307)
 Dober dan, ob kateri uri odide spalnik iz Ljubljane za Pariz (M1308)
 Zdravo, ob kateri uri gre vlak proti Celju (M1309)
 Pelje direktni vlak iz štor v Hrastnik (M1310)
 Pozdravljeni, rad bi vedel, kdaj odide spalni vlak iz Kopra proti Koelnu (M1311)
 Koliko stane vozovnica za direktni vlak z Dunaja za Zurich (M1312)
 Koliko stane povratna vozovnica za zeleni vlak iz Maribora do Ljubljane (M1313)
 Koliko popusta imajo študentje (M1314)
 Rad bi vedel, ali še gre danes kakšen vlak proti Divači (M1315)
 Lep pozdrav, ob kateri uri odide jutri potniški vlak iz Sevnice proti štoram (M1316)
 Ali je še kakšna cenejša varianta (M1317)
 Je možen kak študentski popust (M1318)
 Koliko stane dijaška vozovnica iz Kranja do Ljubljane (M1319)
 Dober večer, prosim, koliko stane tedenska karta iz Ljubljane do Jesenic (M1320)
 Dobro jutro, zanima me, ali odide kak vlak iz Pragerskega proti Kopru (M1321)
 Rad bi izvedel, kdaj bo odpeljal vlak za Budimpešto (M1322)
 Zdravo, ali pelje zeleni vlak s postaje v Mariboru do Kranja (M1323)
 Pozdravljeni, zanima me, kdaj odpelje ekspresni vlak iz Postojne (M1324)
 Koliko zamude bo imel Emona ekspres iz Trsta (M1325)
 Povejte mi prosim, ali bo brzovlak iz Divače zamudil (M1326)
 Ima ekspresni vlak iz Murske Sobote kaj zamude (M1327)
 Je Zeleni vlak iz smeri Pragersko pripeljal z zamudo (M1328)
 Bom za prestopanje porabil veliko časa (M1329)
 Je potrebno iti na drugi peron (M1330)
 Prosim, povejte mi, kakšne so zveze iz Ljubljane za Zagreb (M1331)
 Je potrebno še kje presedati (M1332)
 Kolikšno je doplačilo za prvi razred (M1333)
 Ima pogosto zamudo (M1334)
 Mi lahko rezervirate sedež (M1335)
 Koliko stane karta (M1336)
 Pelje vlak proti Ormožu čez dve uri (M1337)
 Obstaja kaka ugodnejša relacija (M1338)
 Obstaja popust za invalide (M1339)
 Koliko stane vlak iz Divače do Zagorja (M1340)
 Zdravo, koliko stane mesečna vozovnica za ekspresni vlak iz Novega mesta do Sevnice (M1341)
 Lep pozdrav, prosim, ali mi lahko poveste, ob kateri uri bo odpeljal kak vlak iz Laškega (M1342)
 Ali je odpeljal kak vlak za Trbovlje (M1343)
 Dobro jutro, pelje kak ekspresni vlak iz Maribora na Dunaj (M1344)
 Želim izvedeti, ob kateri uri pelje ekspresni vlak iz Maribora proti Ljubljani (M1345)
 Dober večer, rad bi vedel, koliko zamude bo imel brzovlak iz Zagreba (M1346)
 Dober dan, je zeleni vlak iz Zagorja pozen (M1347)
 Dober večer, rad bi vedel, ali ima potniški vlak iz Laškega kaj zamude (M1348)
 Dober dan, rad bi vedel, ali bo potniški vlak iz Dobove prispel z zamudo (M1349)
 Dober večer, ali ima vlak iz Ljubljane kaj zamude (M1350)
 Je treba večkrat prestopiti (M1351)
 Povejte mi, prosim, ali je vlak za Pariz direkten (M1352)

Moram še kje prestopati (M1353)
Rezervirajte mi sedež v oddelku za nekadilce (M1354)
Koliko stane vozovnica za prvi razred (M1355)
Ob petnajst trideset (M1356)

8.5 Corpus MMC 3A

Kdaj odpelje vlak iz Gradca (M1357)
Ob kateri uri prispe vlak z Dunaja (M1358)
V vagonu s kupeji (M1359)
Popoldan (M1360)
Ni druge možnosti (M1361)
Koliko stane povratna karta (M1362)
Lahko noč (M1363)
Jutri (M1364)
Koliko zamude bo imel potniški vlak iz Hrastnika, ki mora priti ob 10:30 (M1365)
Dober dan, zanima me, ali bo brzi vlak iz Zuricha imel zamudo (M1366)
S kolikšno zamudo bo pripeljal Intercity iz Postojne (M1367)
Ob kateri uri prispe vlak iz Prage v Ljubljano (M1368)
Dober večer, zanima me, ob kateri uri prispe direktni vlak na postajo v Postojno (M1369)
Mi lahko poveste, ob kateri uri je potem vrnitev (M1370)
Kako dolgo potuje spalni vlak iz Amsterdama v Muenchen (M1371)
Na kateri tir prispe spalni vlak iz Amsterdama (M1372)
Dober večer, prosim, ali pripelje vlak iz Ormoža čez eno uro (M1373)
Mi lahko poveste, ob kateri uri pelje naslednji vlak za Hrastnik (M1374)
Kdaj pelje naslednji ekspresni vlak za Zidani Most (M1375)
Do katere starosti imajo otroci še popust (M1376)
Koliko stane karta za ležalnik na direktnem vlaku iz Ljubljane do Muenchena (M1377)
Ali obstaja tudi direktni vlak (M1378)
Kdaj (M1379)
Lahko kupim vozovnico tudi na vlaku (M1380)
Zjutraj (M1381)
Kaj pa drugi razred (M1382)
Koliko prtljage lahko imam s seboj (M1383)
Najlepša hvala (M1384)
Poljutrišnjem (M1385)
Rad bi vedel, koliko zamude bo imel ekspresni vlak iz smeri Berlin (M1386)
Dober dan, mi lahko poveste, prosim, koliko zamude bo imel Savinja ekspres iz Celja (M1387)
Bo potniški vlak iz štor kaj zamujen (M1388)
Dober večer, rad bi vedel, za koliko zamuja ekspresni vlak iz Trsta (M1389)
Dober dan, rad bi izvedel, kdaj pripelje kak vlak z Jesenic (M1390)
Mi lahko poveste, kdaj se je mogoče vrniti (M1391)
Dober dan, koliko časa traja potovanje z brzovlakom iz Maribora do Kopra (M1392)
Pozdravljeni, na katerem tiru stoji vlak iz smeri Lendava (M1393)
Dobro jutro, prosim, na kateri peron bo prišel brzovlak iz Kopra (M1394)
Dobro jutro, prosim, ali vlak iz Murske Sobote danes popoldne pride (M1395)
Rad bi vedel, kdaj gre naslednji vlak v smeri za Ormož (M1396)
Lahko plačam s čeki (M1397)
Koliko mora biti potnikov, da lahko dobimo popust za skupino (M1398)
Pozdravljeni, zanima me cena enosmerne vozovnice direktnega vlaka za Krško (M1399)
Kolikšna je razlika v ceni med ležalnikom in spalnikom (M1400)
Zanima me cena vozovnice za potniški vlak iz Dobove do Zidanega Mosta (M1401)
Povejte mi ceno vozovnice za prvi razred ekspresnega vlaka na relaciji Maribor-Koper (M1402)
Lahko prosim izvem, kdaj odpelje naslednji poslovni vlak za Ljubljano (M1403)
Dober dan, lahko izvem, ali bo šel brzovlak za Dunaj pred tretjo uro zjutraj (M1404)
Ali pride kakšen vlak iz smeri Divača pred osmo uro zvečer (M1405)
Na katerem tiru stoji potniški vlak z Dunaja (M1406)
Dober dan, zanima me čas trajanja vožnje z direktnim vlakom iz Trsta v Pariz (M1407)
Zanima me, ob kateri uri se vlak vrača (M1408)
Pozdravljeni, prosim, kdaj bo pripeljal zeleni vlak iz Ljubljane (M1409)
Koliko je zakasnil potniški vlak iz smeri Laško (M1410)
Dober dan, povejte mi, s kolikšno zamudo bo prišel vlak iz Zagreba (M1411)

Želim zvedeti, koliko zamude je imel Emona ekspres, ki bi moral priti pred eno uro (M1412)
 Dober dan, mi lahko poveste, koliko zamude ima brzi vlak iz Lipnice (M1413)
 Ta mi ne ustreza (M1414)
 Koliko (M1415)
 Dobro jutro (M1416)
 Je na vlaku mogoče spati (M1417)
 Ja, zapisal sem (M1418)
 Ali obstaja ugodna zveza med Mursko Soboto in Kranjem (M1419)
 Ob oknu (M1420)
 Kdaj približno bo vlak pripeljal (M1421)
 Ob katerih dnevih je vožnja z vlakom po Franciji cenejša (M1422)
 Prosim, kolikšna je cena povratne vozovnice za prvi razred na relaciji Celje-Divača (M1423)
 Kolikšna je cena vozovnice za brzovlak do Postojne (M1424)
 Ob kateri uri bo odpeljal naslednji vlak proti Jesenicam (M1425)
 Dober večer, rad bi izvedel, ali je še kak vlak za Kranj pred triindvajseto uro (M1426)
 Pozdravljeni, ali prispe brzovlak iz Maribora pred deseto uro (M1427)
 Dober večer, na katerem peronu bo ustavil zeleni vlak (M1428)
 Dobro jutro, prosim, koliko časa traja vožnja s potniškim vlakom iz Kopra za Lipnico (M1429)
 Kdaj je povratek (M1430)
 Dobro jutro, kdaj bo prispel ekspresni vlak iz Kopra v Celje (M1431)
 Dober dan, za koliko zamuja Savinja ekspres (M1432)
 Dober večer, s kolikšno zamudo je pripeljal Zeleni vlak iz Dobove (M1433)
 Dober dan, ali bo poslovni vlak iz Muenchena imel zamudo (M1434)
 Mi lahko poveste, prosim, koliko zamude je imel potniški vlak iz Hrastnika (M1435)
 Hvala (M1436)
 Samo trenutek (M1437)
 Dober dan (M1438)
 Ima vlak tudi spalni vagon (M1439)
 Po vozovnice pridem jutri zjutraj (M1440)
 Bo vlak imel dodatne vagone (M1441)
 Do katere ure delate (M1442)
 Je to Intercity vlak (M1443)

8.6 Corpus MMC 3B

Imam popust, ker sem stara več kot šestdeset let (M1444)
 Imam kaj popusta, če potujem ob koncu tedna (M1445)
 Nudite popust za potovanja ob koncu tedna (M1446)
 Kdaj pelje naslednji brzovlak za Zagreb (M1447)
 Zdravo, zanima me, kdaj bo šel pojutrišnjem vlak Pohorje ekspres za Koper (M1448)
 Bo prišel zeleni vlak iz Ljubljane pred tretjo uro popoldne (M1449)
 Pozdravljeni, kje bo stal vlak iz Zuricha (M1450)
 Dobro jutro, prosim, kako dolgo vozi Intercity vlak iz Poljčan do Celja (M1451)
 Kdaj pa se vrne (M1452)
 Ob kateri uri pripelje vlak Pohorje ekspres v Zidani Most (M1453)
 Povejte mi, prosim, za koliko bo zamujen vlak iz Pariza, ki mora priti popoldan (M1454)
 Zanima me, s kolikšno zamudo pride poslovni vlak iz Dunaja (M1455)
 Dober dan, rad bi vedel, ali je vlak iz Ljubljane morda imel kaj zamude (M1456)
 Dober dan, koliko zamude bo imel poslovni vlak iz Berlina (M1457)
 Hvala lepa (M1458)
 Počakal bom (M1459)
 Dober večer (M1460)
 Se je vozovnica podražila (M1461)
 Je dovolj zgodaj (M1462)
 Kolikokrat bom moral prestopiti (M1463)
 Imate odprto tudi ob nedeljah (M1464)
 Ima vlak spalni vagon (M1465)
 Kako je s popusti za potovanje ob koncu tedna (M1466)
 Kolikšna je cena spalnika iz Bonna do Amsterdama (M1467)
 Kolikšen je popust za upokoјence (M1468)
 Rad bi izvedel, ob kateri uri bo šel kak vlak za Gradec (M1469)
 Kdaj pelje naslednji vlak za čakovec (M1470)

Dober dan, rad bi vedel, ali bo spalni vlak iz Bruslja jutri zvečer pripeljal (M1471)
 Zdravo, na kateri tir pride vlak iz Prage (M1472)
 Dober večer, zanima me, koliko časa potuje poslovni vlak iz Pragerskega do Zidanega Mosta (M1473)
 Zanima me, ob kateri uri se lahko vrnem (M1474)
 Dobro jutro, kdaj pripelje spalni vlak iz Pariza na Dunaj (M1475)
 Dober dan, kdaj pride kak brzovlak iz Bruslja (M1476)
 Pozdravljeni, s kolikšno zamudo je prispel vlak iz Amsterdama (M1477)
 Bi lahko zvedel, ali ima potniški vlak iz Jesenic kaj zamude (M1478)
 Dober večer, mi lahko poveste, koliko zamude je imel vlak iz Amsterdama, ki bi moral priti zjutraj (M1479)
 Srečno (M1480)
 Danes (M1481)
 Na svidenje (M1482)
 Koliko popusta imam, če potujem v nedeljo (M1483)
 Raje bi potoval pozneje (M1484)
 V dopoldanskem času (M1485)
 Rezervacijo za dva sedeža, prosim (M1486)
 Kolikšna je cena spalnika (M1487)
 Kje vse moram prestopiti (M1488)
 Sem že zamudil (M1489)
 Stoji vlak v Divači (M1490)
 Zvečer (M1491)
 Je že prispel (M1492)
 Ima vlak poseben prostor za prtljago (M1493)
 Samo še nekaj (M1494)
 Samo, da si zapišem (M1495)
 Dober dan, koliko zamude ima poslovni vlak iz Poljčan (M1496)
 Koliko zamude bo imel Pohorje ekspres iz Kopra (M1497)
 Dober večer, ali je IC vlak iz Murske Sobote imel zamudo (M1498)
 Zanima me, koliko zamuja Pohorje ekspres (M1499)
 Zanima me, kdaj bo prišel ekspresni vlak iz Maribora (M1500)
 Bi lahko izvedel, kdaj je povratek (M1501)
 Kako dolgo traja vožnja s potniškim vlakom od Dobove do Lendave (M1502)
 Pozdravljeni, zanima me, kje je stal brzovlak iz Maribora (M1503)
 Kam bo pripeljal zeleni vlak iz Ljubljane (M1504)
 Lep pozdrav, prosim, ali pelje kak vlak za Lipnico pred trinajsto uro (M1505)
 Kdaj bo odpeljal naslednji brzovlak proti Trstu (M1506)
 Lahko plačam s kreditno kartico (M1507)
 Koliko popusta imajo invalidi (M1508)
 Kolikšna je cena vozovnice za drugi razred (M1509)
 Koliko zamude ima vlak iz Pariza, ki bi moral prispeti ob dvajset petintrideset (M1510)
 Drugi razred (M1511)
 Kdaj prispem v Postojno (M1512)
 Kdaj pravite (M1513)
 Raje potujem ponoči (M1514)
 Ob kateri uri (M1515)
 Poveste še enkrat (M1516)
 Dober večer, zanima me, koliko zamude je imel vlak iz Prage (M1517)
 Dober večer, koliko zamude je imel Intercity iz Koelna (M1518)
 S kolikšno zamudo bo prispel Savinja ekspres (M1519)
 Pozdravljeni, povejte mi prosim, za koliko bo kasnil vlak iz Prage, ki bi moral priti dopoldan (M1520)
 Pozdravljeni, prosim, kdaj bo prišel vlak iz Pariza (M1521)
 Rad bi vedel, kdaj se lahko vrnem (M1522)
 Dober dan, rad bi vedel, koliko traja potovanje z zelenim vlakom iz Celja v Ljubljano (M1523)
 Dober dan, rad bi vedel, na katerem peronu bo stal poslovni vlak iz Kopra (M1524)
 Rad bi vedel na kateri peron bo prispel potniški vlak iz Gradca (M1525)
 Dober dan, zanima me, ali odide brzovlak za Koper jutri pred enajsto uro (M1526)
 Kdaj odpelje naslednji spalni vlak proti Berlinu (M1527)
 Lahko dobim rezervacijo sedeža za vlak iz Ljubljane do Dunaja (M1528)
 Kolikšna je razlika v ceni vozovnice zelenega vlaka in Pohorje ekspresa iz Hrastnika do Ljubljane (M1529)
 Koliko stane drugi razred (M1530)

9. Set of used sentences for corpus LINGUA

9.1 Corpus Lingua 1A

Pravijo, da je zelo socialen (L0001)
Se bom kar sede umil (L0002)
Imeli so neenakopraven položaj (L0003)
Znal je zelo lepo tkati (L0004)
Moral se je naučiti shajati tudi brez nje (L0005)
Lahko se je tako vedla, saj je bila županova hči (L0006)
Ustavil se je pred njo (L0007)
Hrana se prične kvariti šele po dveh tednih (L0008)
Želijo si snidenja tudi prihodnje leto (L0009)
Noč pokriva goro in log (L0010)
Vzemi, vse vzemi (L0011)
Gre za zanimiv pravni primer (L0012)
Trdo delo med azijskimi tigri ne pomeni le besedne igre (L0013)
Nekaterih ključnih premikov se brez zakonskega žiga oziroma parlamentarne privolitve tudi ne da izvesti (L0014)
Vem, da je bilo sredino jutro, ko smo šli (L0015)
Optimistična naj bi bila predvsem srečevanja pogajalskih skupin (L0016)
Bistvena značilnost je torej opiranje na notranje moči v podjetju (L0017)
Zahtevana sredstva in finančni in poslovni rezultati podjetij pa so odvisni od finančne sposobnosti podjetij (L0018)
Obstoječe delničarje in morebitne nove prepričujejo, naj kupijo nove bančne vrednostne papirje (L0019)
Kdo so ključni ljudje (L0020)
Dobro so se zavedali možnosti zaposlitve v tujini, v isti sapi pa so rekli, da je nihče ne bo iskal (L0021)
Še svatje in pogrebci ne morejo čez most, ne da bi se ustavili na vratih (L0022)
Malo hoje mu ne bi škodilo, saj tako ni imel nič drugega početi (L0023)
Vedeti moramo, kako se otrok določene starosti igra, kaj ga zanima, kaj zmore telesno in duševno (L0024)
Je reven kot cerkvena miš (L0025)
V mestu vsak večer odmevata pesem in smeh (L0026)
Dodali so pet novih interaktivnosti, ki zahtevajo sodelovanje uporabnikov (L0027)
Polna interaktivnost zahteva vračanje in spreminjanje prejšnjih odločitev (L0028)
Vprašal je pacientko, kje jo boli (L0029)
Sedem jih bo imela (L0030)
Hotel je pooblastiti svoje otroke (L0031)
Pričelo ga je žgati po vsem telesu (L0032)
Kar Špela so ji rekli (L0033)
Pričel je zbirati vse kar mu je prišlo pod roke (L0034)
Ne brij norca iz mene (L0035)
Pričel ga je hvaliti, da nam je že vsem presedalo (L0036)
Kupil je tri krofe (L0037)
Ves truden si (L0038)
Včasih je bilo vse drugače (L0039)
V revizijskem postopku je bilo namreč ugotovljeno, da so v podjetju prodajali delnice pod nominalno vrednostjo (L0040)
Ne vladamo gospodarstvu, ampak smo njegovi služabniki (L0041)
Rečeno mi je bilo, da se bodo lastninski odnosi med ustanovitelji uredili v dveh tednih (L0042)
Rekli so nam, da bomo tam na varnem (L0043)
Kar nekaj ur so se pogajali in kupčkali možne kandidate (L0044)
Namen te razprodaje sredstev je pridobiti dodatna sredstva (L0045)
Lahko se odločimo, da ga bomo pokrili z lastnim kapitalom ali dodatnim dolgom (L0046)
Obstaja tudi vedrejši scenarij (L0047)
Narediti bi morali klub, v katerem bi politika in gospodarstvo delala z roko v roki (L0048)
Takšen odgovor je bil osvežitev in menim, da imajo prav (L0049)
Nasmehnil se je, nato se je skril za svoj nasmeh kot za čarovnikovo ruto in izginil (L0050)
Sobo je grela velika pločevinasta peč, v kateri je gorela žagovina (L0051)
Preden se odločimo za nakup, razmislimo, kaj vemo o otroku (L0052)
V hiši je polno miši (L0053)
Kam gre (L0054)
Med preferencami je tudi nastavitev barv na zaslonu (L0055)
Prijeten, zabaven, tudi duhovit programček (L0056)

Zanimalo me je, koliko se lahko zanesem na slovnični program (L0057)
 Stvari so vedno cenejše, če jih kupujemo na veliko (L0058)
 Merijo cesto čez reko (L0059)
 Glej, Breda že ve (L0060)
 Ali si sit (L0061)
 Dobiček namreč kujejo z lepimi dekleti, ne pa z dobrimi frizerji (L0062)
 Vsi pisatelji romanov vsaj v glavnem soglašajo, ko opisujejo svet, v katerem živimo (L0063)
 Potem je zmagovalka Atena (L0064)
 Največjo korist bo od lastninjenja namreč imela prav država, ki si je lastninjenje izmislila (L0065)
 Imamo tri podjetja, ki so si med seboj konkurenčna (L0066)
 Seveda obstajajo načini, kako preprečiti prevzem podjetja (L0067)
 Na podlagi teh tabel ugotovimo, kakšno financiranje je najprimernejše za naše podjetje (L0068)
 Vaški vrč si zapomni vse zdravstvene težave plemena (L0069)
 Ne vem zakaj, ampak tako je (L0070)
 Zaradi nastale moralne in materialne škode je zadevo že izročil svojemu odvetniku (L0071)
 Kandidiraj, le kandidiraj (L0072)
 V državi, ki proizvede in izvozi največ nafte na svetu, so skoraj podvojili cene bencina (L0073)
 Videl sem le rob vaze (L0074)
 Tu kupi klobuk (L0075)
 Glej, lepo mesto (L0076)
 Bila je prelepa gmajna (L0077)
 Nepremično je pričel zreti vanjo (L0078)
 Kmet je moral zorati celo njivo (L0079)
 Prijel je za steklo in se cel porezal (L0080)
 Na stara leta je postal ves zguban (L0081)
 Ptica je zapela v ranem jutru (L0082)
 Ta primer je neenakozložen (L0083)
 Odločila se je, da bo ubila dve muhi na en mah (L0084)
 Teksti velikih avtorjev pravzaprav diktirajo jezikovne norme, zato bi morali biti najmanj sporni (L0085)
 Diske odlikuje vmesnik za pregledovanje vsebine in instalacijo datotek (L0086)
 Medved je še posebej zvečer huda zver (L0087)
 Jejte to svežo jed (L0088)
 Začni, vsi začnite (L0089)
 Da, med tujci se še najdejo, ki ne vedo veliko o naših prečudovitih vinogradih in še bolj žlahtnih vinih (L0090)
 človeški um še vedno gleda posamične dežele na globusu v nekakšni zmešnjavi, kot živali in simbole (L0091)
 Pa naj stane kolikor hoče (L0092)
 Vas to ne spominja na trditev, da so za kršitve kriva pravila in ne tisti, ki jih krši (L0093)
 Razlike med bankami so vse prej kot očitne, vsaj pri dajanju posojil (L0094)
 Če vrednost delnic ni podcenjena, podjetje torej ni vablljivo za prevzem (L0095)
 Vse naštetu velja pod pogojem, da so dobički od poslovanja večji kot stroški kapitala (L0096)
 Vlada da vklesati zakone v bronasto ploščo (L0097)
 Ob reki so se vedno ustavljali popotniki (L0098)
 To je za sedaj vse (L0099)
 Gostoval je v Italiji, Nemčiji, Romuniji, češki, na Norveškem in Madžarskem (L0100)
 Načrtuje tudi lastninjenje državne lastnine (L0101)
 Ob fantu gre pes (L0102)
 Ovčar je sedel na travi (L0103)
 Veselil se je jutrišnjega dne (L0104)
 Smiliti se mu je pričel šele, ko je bilo že prepozno (L0105)
 Pričelo je tleti na več mestih hkrati (L0106)
 Brenčal je kot čmrlj (L0107)
 Skoraj bi ga bil pahnili v globel (L0108)
 V Gdansk odpotuješ (L0109)
 Grobo je brčnil psico (L0110)
 Le pootročiti se ne pusti (L0111)
 Želel se je utopiti (L0112)

9.2 Corpus Lingua 1B

Ne nauči nas vsega, nas pa nauči marsičesa (L0113)

Z robnimi pisavami je moč izdelati nešteto robov (L0114)
 Metki je vzela pravo mero (L0115)
 Mejni primeri so najtežji (L0116)
 Na veji sedi ptič ob ptiču (L0117)
 Poslej bo libijski denar v tujini zamrznjen, začel pa je veljati tudi embargo na nafto in njene derivate (L0118)
 Ravnina je bila pod večernim soncem tako abstraktna, kakor da jo gledam v sanjah (L0119)
 Pomembno je, da kljub komaj posušenemu črnilu na podpisnem sporazumu o sodelovanju le-to že teče (L0120)
 Je to količina denarja, porabljenega za oglaševanje (L0121)
 Če bančniki niso dovolj radodarni, potem gredo njihove premožnejše stranke h konkurenčni banki (L0122)
 Če se delničarji prepričajo o pravilnosti potez vodstva, potem ne bodo želeli prodajati delnic podjetja po nizki ceni (L0123)
 Ko je trg nelikviden, se običajno trguje s samo majhnimi količinami vrednostnih papirjev (L0124)
 Za vse svete knjige je znano, da baje veljajo dalj časa (L0125)
 Certifikati niso prenosljivi (L0126)
 Po klicanju desetih zasebnikov z istim priimkom se je izkazalo, da nihče med njimi ni lastnik portreta (L0127)
 Zbor ima trenutno tri vsebinsko različne programe in goji vse vrste pesmi (L0128)
 Izvedenci že nekaj časa zahtevajo tesnejšo povezavo trgov, da bi povečali trgovino med arabskimi državami (L0129)
 Z zobmi so vedno težave (L0130)
 Na pol gol je hotel hoditi (L0131)
 Prinesi nam še rdečega (L0132)
 Zakaj si napravil takšen zmazek (L0133)
 Dleto boš našel čisto na dnu (L0134)
 Na polju je videl rasti hmelj (L0135)
 Scela se ga moraš lotiti (L0136)
 Podano čtivo ga je pričelo zanimati (L0137)
 Drobnim snežinkam pravijo babje pšeno (L0138)
 Bil jim je neenak (L0139)
 Kot za šalo je ukrivil debelo palico (L0140)
 Zamere programu so malenkostne (L0141)
 Uporabniški vmesnik, ki je pri Microsoftovih programih odličen, so izboljšali in ga dopolnili z nekaj možnostmi (L0142)
 Celo leto je delal z lesom (L0143)
 Katera je vzela Večer (L0144)
 Domači sir je boljši od kupljenega sira (L0145)
 Kdor misli, da se bodo zaradi neurejenih razmer v stanovanjskem gospodarstvu začele podirati hiše, se moti (L0146)
 Tujec, ki se je usedel k točilni mizi, se je široko zasmel (L0147)
 Začeli smo z majhnimi koraki in ob vztrajnosti bo čas prinesel svoje (L0148)
 Je morda merilo, katere družbe se ljudje najprej spomnijo (L0149)
 Prav to je nekakšno protislovje (L0150)
 Prvi pogoj za uspešno obrambo pred prevzemom podjetja je prepoznati strategijo podjetja, ki si želi prilastiti naše podjetje (L0151)
 V takšnem trenutku ima tudi zelo majhna transakcija zelo velik vpliv na cene vrednostnih papirjev (L0152)
 Prva metoda je metoda preobrata (L0153)
 Ali bo to izvedel prek pooblaščenja (L0154)
 Velik problem predstavljajo tudi parkirišča in prenočitvene zmogljivosti (L0155)
 Jeseni bi radi ustanovili še otroški in mladinski pevski zbor (L0156)
 Vsaj na začetku naj bi bile informacije iz elektronske baze podatkov brezplačne (L0157)
 To je pa tudi v interesu delničarjev je še pojasnil (L0158)
 No, govori (L0159)
 Saj vas čakava, Nana (L0160)
 Na tvalo ga je posadil (L0161)
 Hlod ga je pokopal pod seboj (L0162)
 Ves čas je prebiral dnevni časopis (L0163)
 Ščasoma je nekoliko popustil pritiskom (L0164)
 Spanje ti bo koristilo (L0165)
 Nehaj me phati naprej (L0166)
 Kmalu nato se je pooblašilo (L0167)

Poslušali so uleknjeni na tleh (L0168)
 Celuloid je osnova za izdelavo filmskega traku (L0169)
 Zahteval je imenski seznam (L0170)
 Bili so si needini (L0171)
 Ksaver, da ti je ime (L0172)
 Števec se je neusmiljeno vrtel (L0173)
 Zdeti se mu je pričelo, da ima privide (L0174)
 Letos pa me ne bo na Cres (L0175)
 Odprl se je prelep cvet (L0176)
 Bil je zdrav kot dren (L0177)
 Smrt ima tudi zate prt (L0178)
 Popelji popotnika spat (L0179)
 Natečaja so se udeležila štiri podjetja (L0180)
 Ljudje menijo, da je težko delo dolgočasno (L0181)
 Lepo me je prosil, naj tega ne storim (L0182)
 Tam smo ostali sedem dni (L0183)
 Proizvodni načrti bodo odvisni tudi od oskrbe z materialom (L0184)
 Skratka, lahko rečemo, da v tem trenutku slovenski trg ni primeren za managerske odkupe (L0185)
 Naslednja stopnja je določitev pravilne kapitalske strukture (L0186)
 Kaj pa, če je šlo samo za vljudnostno frazo (L0187)
 To je čisto drug fokus (L0188)
 Zapri se v hotelsko sobo in glej televizijo (L0189)
 Dan je res, ker pa ne vem, kateri, jutrišnji ali včerajšnji, zame ni veliko jasnejši kot noč (L0190)
 Najbolj pa se je mnenje javnosti spremenilo glede novinarjev (L0191)
 Ob računalniku otrok ne more ostati pasiven, saj stroj vedno potrpežljivo čaka na odziv oziroma na ukaz (L0192)
 Fantič Martin sili k teti Mili (L0193)
 Onečedel je njeno lepo obleko (L0194)
 Čim nižja je številka, tem zahtevnejše je branje (L0195)
 Odpre se zaslon z dvema okencema (L0196)
 Želel je videti materialen dokaz (L0197)
 Ni mu imel kaj poočitati (L0198)
 Solato moraš še pooljiti (L0199)
 Moral boš bdeti vso noč (L0200)
 Škarje so rezale kar same (L0201)
 Nehaj ždeti v kotu in stopi semkaj (L0202)
 Nehaj že srebati to juho (L0203)
 Še sam ne vem, kaj ga je pričelo gnati (L0204)
 Miš lovi miši (L0205)
 Padel je čez brv (L0206)
 Zob me boli (L0207)
 Zanj se je ogrel zaradi strokovnosti, izkušenosti in zaradi zanimive cene njenih storitev (L0208)
 Tokrat je bilo še posebej slovestno (L0209)
 Da se bodo stvari kmalu uredile, je dejal, in da naj počakam vsaj do centrove druge obletnice (L0210)
 Doma smo imeli najraje zimo (L0211)
 Ja, v ledeni dobi so praljudi razveseljevali pripovedovalci zgodb (L0212)
 Torej prvi pogoj so nizke cene delnic na trgu (L0213)
 Pri analiziranju prevzemov si pomagamo tudi s preverjanjem finančne moči podjetja, ki namerava prevzeti podjetje (L0214)
 Presenetili sta jih homogenost in izobrazba delovne sile v Sloveniji (L0215)
 Ko pravim, da Slovenci izgubljam v tekmi s časom, mislim resno (L0216)
 Velikanska stara gostilna v obliki pravokotnika je bila polna kakor panj (L0217)
 Prinesla mi je lep kos kruha in kepo sira, zavito v lapuhov list (L0218)
 Vsak drugi slepi Tajvanec dela kot maser (L0219)
 Kupil je več sit (L0220)
 Tista siva nit je čisto zanič (L0221)
 ZDA imajo veliko prednost vsekakor zaradi angleščine, visoke ravni računalniške pismenosti in gostih linijskih servisov (L0222)
 V poslovnem in tehničnem pisanju se ogibajte odstavkov z enim stavkom (L0223)
 Spodaj so običajne bližnjice (L0224)

9.3 Corpus Lingua 2A

Reci Cilki, naj ne cenca s cvičkom (L0225)
Gledališču je mogoče verjeti, če je mogoče verjeti gledališki kritiki (L0226)
Vso vino v kleti se je razlilo (L0227)
Nehaj se no že cmeriti (L0228)
Postal je bled od strahu (L0229)
Mnogo jih je bilo, sedaj pa ni nikogar več (L0230)
Tega mi pa že ne bodo vcepili (L0231)
Tega si nikakor ne pustim vzeti (L0232)
Moral je vstaviti umetni kolk (L0233)
Nameraval je zbrati čim več prostovoljcev (L0234)
Tnalo je padlo na dno (L0235)
Dota se hitro izmota (L0236)
Ničesar nisem prav dojel, si ničesar zapomnil (L0237)
Naša vlada priporoča čim poznejše sklepanje zakonskih zvez (L0238)
Mi je sploh kaj zares hudega (L0239)
Začuda, takšnih je menda malo (L0240)
Pluta je naraven organski material, ki se že dolgo uporablja za toplotno izolacijo zgradb (L0241)
Smučišče je osvetljeno, tako da je možna tudi nočna smuka, proge pa so redno vzdrževane (L0242)
Programerji smo nagnjeni k temu, da iščemo čarovnijo, ki nas bo odrešila vseh težav (L0243)
Na širšem območju Maribora zato iz prometa izločajo tovorna vozila s prikolicami in vlačilce, za preostala pa je obvezna zimska oprema (L0244)
Po njegovem mnenju bi takšen koncept pomenil postopno ukinjanje lokacije štrore in prenos tehnologije (L0245)
Njemu in ženi nesi, ne meni (L0246)
Njemu je bil ded vse na svetu (L0247)
Mraz jo je grizel v obraz (L0248)
Tu kupi svež kruh (L0249)
Jud jo je odkuril na jug (L0250)
V oglatih oklepajih so v tekstu zapisani predlogi popravkov (L0251)
Nič ne damo na občutke (L0252)
Stric gre h Kocbeku (L0253)
Gledališče v zadnjih štirih letih ni obdelalo nobene zanimive domače zgodbe (L0254)
Rekel mu je, da je suh kakor glista (L0255)
Šmohor se je jezno obrnil (L0256)
Si samo člen v verigi in nič drugega (L0257)
Vodo mu je hotel vlti za ovratnik (L0258)
Moral mu je všiti vse krpe (L0259)
Vsiliti je hotel svoje mnenje (L0260)
Nisem se smel všteti (L0261)
V tej hiši je moral zblazneti vsak, ki je stopil vanjo (L0262)
Tla so spolska (L0263)
Strah se plazi po svetu (L0264)
Je bil park zelen, bi lahko bil takšen na začetku zime (L0265)
Vseeno moram vzeti ponujeno posetnico (L0266)
Morda lahko strah preženemo samo z zastraševanjem (L0267)
Zapeče me in zato postanem žalosten še bolj (L0268)
Če drevesa pravilno luščijo, jih ne poškodujejo, tudi predelava ne zahteva velike rabe energije (L0269)
V drugem delu je potekal pogovor o aktualnih in perečih problemih današnjega časa, kot so droge in druge oblike zasvojenosti (L0270)
Knjiga je namenjena tistim, ki že znajo kak višji programski jezik (L0271)
Deponija je skoraj prazna in zaloge zadoščajo le še za dvakratno ali trikratno soljenje (L0272)
Večini sogovornikov v regiji grozi, da v naslednjih letih ne bodo več vozili s tovornjaki v tujino (L0273)
Le veseli se je (L0274)
Študent je pregledal nek segment (L0275)
Pozlatil nam je čas (L0276)
Strl je suho brv (L0277)
Šel je res truden (L0278)
Program omogoča, da dodamo nove stile, podobno kot lahko v črkovanju ustvarimo posebne slovarje (L0279)
Pravijo, da so bili še sami presenečeni nad povpraševanjem po tem vmesniku (L0280)
Tatovi mirno živijo med nami, ne da bi se kdo razburjal, ne da bi jim kdorkoli skrivil las na glavi (L0281)

Programska oprema, ki mrežo poganja, je združljiva z Novellovo (L0282)
 Starec ne more premakniti čebel (L0283)
 Dober kup je sklenil (L0284)
 Naglas je vabil pravi naglas (L0285)
 Da, včasih nam kaj da (L0286)
 Peta žena je bila pa že preveč (L0287)
 Imidž je zelo dobro prodajani izdelek, čeprav se običajno sploh ne ve, ne kdo kaj prodaja ne kdo kaj kupuje (L0288)
 Nekoliko ostrejša formulacija v primerjavi z inačico, ki je krožila pred konferenco, ni naključna (L0289)
 Funkcije, ki delajo s strukturo, kličemo podobno, kot običajno dosegamo podatke (L0290)
 Strokovnjaki zagotavljajo, da je strah popolnoma odveč, saj so to divje ptice, ki znajo za razliko od tistih udomačenih v parku same poskrbeti zase (L0291)
 Mlademu sadnemu drevju požre korenine prav do debela, tako da spomladi potegnemo iz zemlje le še oglodan štrcelj (L0292)
 Rad bi videl, kje bo zašlo sonce (L0293)
 Kar poglej se, kakšen si (L0294)
 Zanj niso krive trenutne okoliščine, te so samo zrcalo (L0295)
 Vrnil se bom, da ne zamudim avtobusa (L0296)
 Še nikoli ni slišal takšnega šuma (L0297)
 Napolni škaf z vodo (L0298)
 Skočila je nadenj in mu hotela spraskati obraz (L0299)
 Drog je moral skriviti s posebno pripravo (L0300)
 Zanj je bil to prehud vzpon (L0301)
 Včeraj si pa mogel priti (L0302)
 Kmalu nato je muca pričela mjavkati (L0303)
 Kje so sedaj tvoje sanje (L0304)
 K sosedu grem (L0305)
 Od koder si prišel, tja pojdi (L0306)
 Vse to te lahko spravi v depresijo, če nisi pripravljen in si želiš česa drugega (L0307)
 Izrezal mu je slepo črevo (L0308)
 Pomembno vlogo igrajo senzorji (L0309)
 Program nam v svoji bazi poišče prevod, tega pa lahko potem vrinemo v dokument z beležnico (L0310)
 Priden si, je rekel dedek (L0311)
 Ubila je jajce v juho (L0312)
 Pravzaprav takrat tamar tamkaj še ni stal (L0313)
 Brat gre rad jagode brat (L0314)
 Tele krave in še eno tele vam privedem (L0315)
 Simbolni in vedenjski odnosi med podjetjem in njegovimi javnostmi so tako prepleteni, da jih ni mogoče povsem razločiti (L0316)
 V zadnjem desetletju je prevzel vlogo glavnega izpraševalca pri zaključnih izpitih, kar bo nedvomno ostalo v spominu mlajših generacij fizikov (L0317)
 Iz Bagdada poročajo, da so eksplozije povsem razsvetlile nebo nad mestom, spremljalo pa jih je zavijanje siren (L0318)
 Podobe našega mesta si ni mogoče več zamisliti brez elegantnih belih ptic na Dravi, zato te hladne dni Mariborčane skrbi, kako bodo labodi preživeli zimo (L0319)
 Drevo, ki smo ga pred zimo pravilno negovali in zaščitili, nam bo v naslednjem letu povrnilo z obilnim pridelkom (L0320)
 Nisem jezen, privoščim jim (L0321)
 Pošta je takoj za vogalom (L0322)
 Zadovoljenosti ali preizkušnje se pač ne da naročiti kot kosa torte ali stekleničke grenkega sirupa (L0323)
 Kljub šaljivosti ga je nekaj tistega, kar ga žgečka po podplatu jezika, pustilo na cedilu (L0324)
 Dež škrablja po šipi (L0325)
 Vilma nima rada filmov (L0326)
 Rad je vzel v roke vsako stvar, a le malo mu je uspelo (L0327)
 Splesti je nameravala veliko košaro (L0328)
 Vzdlž zidu je žuborel umetni potoček (L0329)
 Vhod je bil zatrpan z avtomobili (L0330)
 Bil je njegov in ne tvoj (L0331)
 Tvoj delež je žal odveč (L0332)
 Jutranji frakelj je bil zanj obvezen (L0333)
 Kmalu je bilo vsem jasno, da so zapluli v napačen fjord (L0334)
 Spet se zavem, da ni dobro začeti in končevati tukaj (L0335)

Tole si pa pozabil šrafirati (L0336)

9.4 Corpus Lingua 2B

Vsak sistem ima svoje prednosti in pomanjkljivosti (L0337)

Programu je priloženih približno 70 vzorcev tolkal in drugih ritemskih elementov (L0338)

Trhla brv (L0339)

Uršinemu vnuku je ime Luka (L0340)

Naša praprababica in vaš prapraded sta bila praprebivalca našega kraja (L0341)

Kaj je spraševal (L0342)

Redi več jelenov (L0343)

Tukaj nastopi končni zaplet: vedenjskih odnosov brez simbolnih sploh ni mogoče graditi (L0344)

Zlasti odmevni so njegovi prispevki, kjer obravnava transportne in dinamične probleme v trdnih snoveh (L0345)

Znanstvenik je zavrtel čas nazaj in iz gole lobanje z vdrtim nosom ustvaril glavo človeka, ki naj bi jo imel pred pet tisoč leti (L0346)

Strokovnjaki zagotavljajo, da je strah popolnoma odveč, saj so to divje ptice, ki znajo za razliko od tistih udomačenih v parku same poskrbeti zase (L0347)

V Sloveniji je po najnovejših strokovnih ugotovitvah po drogi vsaj enkrat posegel že vsak tretji mladostnik (L0348)

Strah je bil votel, okrog kraja pa ga ni bilo nič (L0349)

Če je njim, potem je prav tudi meni (L0350)

Torej metlo in omelo v roke (L0351)

Telefoniral sem ji, dobimo se ob četrtr čez dvanajsto (L0352)

Žužnjati je pričelo ob žimnici že jutraj (L0353)

Iz tega soda ti dotočim (L0354)

Takoj je morala sčvekati sosedi kar ji je bil povedal mož (L0355)

Juho moraš shladiti, preden jo lahko zaužiješ (L0356)

Obrni se na vzhod in pomahaj (L0357)

Vsa vas se je pričela vgrezati (L0358)

Moj ljubi brat me je vznejevoljil (L0359)

Ne smeš se sedaj vmešati (L0360)

Samo čez ta hrib pojdi pa boš tam (L0361)

Ves dvor je pokal od smeha (L0362)

Zaradi sebe fotoaparata zagotovo ne nosim s sabo (L0363)

Pričel je žreti kot pujs (L0364)

Vse kar je potovalo od enega računalnika do drugega, so bile oznake objektov in njihove nove koordinate (L0365)

Ko je prišel na trg, je bil prvi program takšnih zmogljivosti za katerikoli računalnik (L0366)

Sledi možnost, da poženemo le črkovalnik, tako da interaktivno preverimo napake v črkovanju (L0367)

Gruntar je grunt zapustil tujim ljudem (L0368)

Nadvse mu zaupa (L0369)

Če ga ima kaj rada (L0370)

Zet je segel v žep (L0371)

Rekla je vsem, ali ne (L0372)

Naš sogovornik priznava, da so obroki za plačilo elektrike za marsikoga zelo visoki (L0373)

Snežne padavine v zahodni Sloveniji so večinoma prenehale, še vedno pa močno sneži v vzhodni Sloveniji (L0374)

Vse več ljudi se posredno ali neposredno srečuje z računalniki in vedno bolj so kritični glede njihovih storitev (L0375)

Med narkomani je preventivno ravnanje, ki bi jih ščitilo pred možnostjo okužbe z virusom HIV, na izredno nizki ravni (L0376)

Mineralna vlakna delimo glede na surovine, iz katerih jih pridobivamo, na kamena in steklena (L0377)

Dan sicer ni bistro čist, vendar je sončen in topel (L0378)

Pisanje pisem je precej jalov nadomestek (L0379)

Samo ugibam lahko zakaj (L0380)

Žaga žal ne dela (L0381)

Kdor hitro da, dvakrat da (L0382)

Obveza mu je popustila (L0383)

Na hitrem posnetku je videl sveteti cvetlico (L0384)

Motor ni hotel vžgati (L0385)

Pozabil je vključiti štedilnik (L0386)

Zelo težko mu je bilo kaj vbiti v glavo (L0387)
 Iz shrambe prinesi mleko (L0388)
 Po mirnem morju ni težko pluti (L0389)
 Močno ga je moral zvezati (L0390)
 Z vsemi temi predsodki vstopim (L0391)
 Tako kleti pa ni lepo (L0392)
 Če bo časa, skočim še na čaj (L0393)
 Rad bi šel na jug (L0394)
 Nenadoma je slika padla na tla (L0395)
 Vsa kri mu je stopila v obraz (L0396)
 Mati je šla od tod (L0397)
 Ni mu uspelo vnesti nemir (L0398)
 Vživeti bi se moral v njeno vlogo (L0399)
 Stkan je bil iz čiste svile (L0400)
 Izkoristil je vzgon vetra (L0401)
 Ob sončnem vzhodu je v snegu zazijala velika špranja (L0402)
 Dno boljše očisti (L0403)
 Zvezde sijejo visoko (L0404)
 In v predsobi je vse, kar ni nujno imeti v stanovanju (L0405)
 Smejoča se ribica, rižjezobkasta muca (L0406)
 Boš našel pot do svoje sobe (L0407)
 Le na koga naj bom hud, ko pa se je tako lepo smejati (L0408)
 Pritisnil je mraz, zapadel je prvi sneg in razen nekaterih zamudnikov smo z vrtov pobrali vse pridelke (L0409)
 V soboškem invalidskem podjetju dela osemdeset delavcev, od tega jih je polovica invalidov (L0410)
 Programiranje je postopek, v katerem programer zgradi model realnosti (L0411)
 V BiH bo ostal štiri dni, v tem času pa bo skušal ugotoviti, kako bi lahko najučinkoviteje organizirali pomoč tukajšnjim prebivalcem (L0412)
 Na Madžarskem je daleč največ ameriškega kapitala, na drugem mestu je nemški in na tretjem avstrijski (L0413)
 Meni, da o meni vse ve (L0414)
 V prijetni atmosferi so si ogledali bolero (L0415)
 Kraj travnika je pravi kraj (L0416)
 Preusmeriti plug (L0417)
 Rjav steber (L0418)
 Vsi objekti se lahko premikajo in so lahko občutljivi za miško (L0419)
 Na sejmu so prikazovali preliminarno, a že delujočo verzijo (L0420)
 Ljudje živijo v skladu z zgodbami, ki so jih prepričale (L0421)
 Tu pa sem preprosto zato, ker moram biti (L0422)
 Kakor bi se bilo vse zlo združilo v njem (L0423)
 Vse je stavil na znanje (L0424)
 Žlica mu ni bila dovolj velika (L0425)
 Pozabil je vrisati zobnik (L0426)
 Mazilo si morate previdno vtreti v kožo (L0427)
 Na prednji strani jope je bil vtkan njegov grb (L0428)
 Otroci so se prerekli kdo zna najbolje scmariti pločevinko (L0429)
 Pričel je šklepetati z zobmi, tako ga je zeblo (L0430)
 Fraze so veselo frčale po grofovem dvorcu (L0431)
 Še tega šušmarja mi boš spravil v hišo (L0432)
 Brez kril je nevarno letati in brez samote je težko živeti z ljudmi (L0433)
 Od časa do časa se zalomi, brez tega ne gre (L0434)
 Prav ima, naj počne z mano, kar ga je volja (L0435)
 Po cikcakastem stopnišču se povzpnem na nadhod (L0436)
 Voluhar je glodalec, ki lahko povzroči v nasadu veliko škodo (L0437)
 Konferenca o lakoti, po mnenju organizatorjev samih, ne bo prinesla kakih bistvenih sprememb (L0438)
 Pri inicializaciji statično hranjenih spremenljivk lahko nastopajo poljubni izrazi (L0439)
 Precej razpravljavcev je tudi menilo, da zakon pušča vnevar vse druge igralnice v Sloveniji, kar gotovo ni dobro (L0440)
 Predvsem pa se ob vsem skupaj porajajo novi pomisleki, nič manj tehtni od že navedenih (L0441)
 Precej priteče Pepa z vedrom in metlo (L0442)
 Vsega še njemu ne rečem (L0443)
 Na, da prav tako babo najdem (L0444)

Jure je kumare metal na kup (L0445)
Pes tuli v veter (L0446)
Program teče v računalnikih, povezanih v lokalno mrežo (L0447)
Kupčije so jim šle lani zelo dobro (L0448)

9.5 Corpus Lingua 3A

Za vsako ceno moramo streti upor še to noč (L0449)
Zgrabiti je moral z obema rokama, da je premaknil voz (L0450)
Ni je maral vznemiriti s svojim prihodom (L0451)
Samo kljun je odprl pa je bilo po njem (L0452)
Nehaj že zvrčati krivdo name (L0453)
Vstrajati moraš do konca (L0454)
Vratolomna vzvratna vožnja ga je stala celo premoženje (L0455)
Po naključju se je znašlo na enem mestu šest Mojc (L0456)
Črne oči ima in visoko čelo (L0457)
Anka gre po Angelo v Anhovo (L0458)
Simfonija danes žal odpade (L0459)
Sonce se kot vselej sredi dneva peče na razbeljeni plošči (L0460)
Knjiga je obdavčena tako kot mleko (L0461)
Torej najprej, da ne bo očitkov, da se izmikam in da sem bežal v popolnoma napačno smer (L0462)
Novo ne more nikoli razočarati (L0463)
Vsak ptič si svojo vižo poje (L0464)
Ko pa se vrneš, te povabim na večerjo (L0465)
V Združenih državah Amerike se vse bolj uveljavlja praksa, da prinašajo raznašalci časopisov naročnikom tudi knjige (L0466)
S kislino oprano nebo so preletavale različno rožnate papige (L0467)
Prodaja alkohola v normalnih razmerah je na Norveškem zelo strogo urejena, dovoljena samo v posebnih trgovinah (L0468)
Če poslanci slednjih dveh ne bi izvolili, naj bi po drugem predlogu ruški izvršni svet še naprej delal v petčlanski sestavi (L0469)
V gostilni zraven višegrajske carinarnice se je počasi nabralo kar precej popotnikov (L0470)
S hvaležnostjo in spoštovanjem se spominjamo slovenskih teritorialcev (L0471)
Postavi stol tesno ob kol (L0472)
Visoko nad vasjo je stal gotski stolp (L0473)
Stikala so dosegljiva prek posebne odprtine na zadnji strani kartice (L0474)
Testiranje hitrosti skenerjev je pravzaprav zelo enostavno (L0475)
CD ROM pogoni postajajo že skoraj nepogrešljivi del osebnih računalnikov (L0476)
Shraniti je pozabila srečko, ki je zadela (L0477)
Raje je pustil zgniti ves krompir, kot da bi ga dal lačnemu sosedu (L0478)
Vgraditi so morali vse kamne, da je most vzdržal (L0479)
Dlje časa kot je hodil, manj je vedel kje je (L0480)
Knjiga je bila pretežka, da bi jo lahko sam dvignil (L0481)
Vštric sta hodila vso pot (L0482)
Zbralo se je veliko pihalnih godb (L0483)
Bil je že vaje različnih prerokb (L0484)
Ta enačba, v tem primeru ne velja (L0485)
Bela Ljubljana (L0486)
Na vratih razreda visi urnik (L0487)
Z dlanjo si naredim senco za oči in podaljšam korake (L0488)
Moral sem nekaj storiti (L0489)
Vsaj kolo gre tja, kamor hočem (L0490)
Če so mi napoti, se jim pač ognem (L0491)
Šaljivec brez šal (L0492)
Je bilo dovolj denarja (L0493)
Berejo ga že milijoni (L0494)
Tam višje so veslale tudi fantastične brole in ko smo nenadoma zavili na desno, smo se znašli v strapani jati kot mačke orjaških, sejestih netopirjev (L0495)
Tako izjemen nagradni sklad je bil sprva sporen, zdaj pa nihče več ne obregne nanj (L0496)
Sicer pa so dravski labodi že tako osvojili srca mladih in starejših obiskovalcev dravskega obrežja (L0497)
Pravijo, da je to velik čuden oder, na katerem sleherni nekaj časa igra svojo vlogo (L0498)
Spekel je dober šarkelj (L0499)

Na grob so nasuli grob pesek (L0500)
 Lahko jih vidimo, slišimo in morda celo čutimo (L0501)
 Moč sodobnih baz podatkov leži v povezavah in poizvedbah (L0502)
 Težave so nastopile le takrat, ko je bila poskenirana slika tako velika, da je računalnik ni mogel spraviti v hitri pomnilnik (L0503)
 Delo je že v polnem teku (L0504)
 Skeniranje lahko poteka na več načinov (L0505)
 Za nadaljnje delo je treba odpreti bazo podatkov (L0506)
 Števec je numerični podatek, ki se mu vrednost avtomatično povečuje in je zaradi tega enolično določen (L0507)
 Ponoči je čul sovo (L0508)
 Kolo je vozilo, kolo je ples (L0509)
 Čevelj mu je bil tesen (L0510)
 Mestna straža bi morala bolj učinkovito delati glede na svoja pooblastila (L0511)
 Pravi ljubitelj živali bo namesto nakupa rodovniškega psa raje rešil izgubljenega oziroma zavrženega psa (L0512)
 To nalogo je galerija še posebej uresničila tudi s sedanjo Adamljevo razstavo (L0513)
 K uspešni predstavitvi je prispevala tudi domiselna postavitev razstavnega prostora in pa animacijski nastopi plesalke, ki so očarale številne obiskovalce sejma (L0514)
 Le kaj jih bom božal, treba jih je trdo udariti (L0515)
 Ti bi to moral razumeti (L0516)
 Priznati moram, da na ta vprašanja ne vem odgovorov (L0517)
 Nevpadljivo se sprehodim do v kamen izklesanega ovinka, potem pa po ozki poti stečem s sedla (L0518)
 Ne sprašujte me, dovolj sem že slišal teh vprašanj (L0519)
 Zakon je nastajal in se spreminjal (L0520)
 Torej kaj bomo naročili (L0521)
 Adapter priključi, pa bo delalo (L0522)
 Samo hihitanje jih je (L0523)
 Dal vam bom recept za zdravilo, ki vam bo lajšalo bolečino (L0524)
 Si nasekal dovolj tresk (L0525)
 To leto smo ustanovili pet novih pošt (L0526)
 Zbljuval je vse kar je zaužil (L0527)
 Zvleci ga s ceste, da ne bo oviral prometa (L0528)
 Ob predvajanih prizorih je pričel krčevito bljuvati (L0529)
 Veliko dodatkov moram vgnesti v testo, da bo kruh krhek in mehak (L0530)
 Samo vprašati bi ga bilo treba pa si ga ni nihče upal (L0531)
 Moral se je zdreti, da si je olajšal dušo (L0532)
 Skener vmesne točke določi z interpolacijo (L0533)
 Eden prvih korakov je izdelava tabel oziroma priklop na zunanje tabele (L0534)
 Najvišja raven podatkov je baza podatkov, v kateri so združene tabele (L0535)
 Skočil je čez plot na betonsko ploščo (L0536)
 Nosil je mojo novo jopo (L0537)
 Čez mrliški obraz so pokrili črn prt (L0538)
 Vlada bo uveljavljanje prometnega režima v modri coni tudi vnaprej skrbno spremljala (L0539)
 Seveda bomo s strankami ugotavljali le neupravičenost ali upravičenost izdaje lastniškega certifikata in zakaj ga stranka ni prejela (L0540)
 Vem da sodim v svetovni vrh, vendar je to treba dokazati na tekmi (L0541)
 Vsak dan en srečen Slovenec več (L0542)
 Še najmanj pa takšnih, ki bi bili sprejemljivi za prizadete skupine ljudi (L0543)
 Čas lahko mirno zaokroživa na štiri ure (L0544)
 Ne bodite smešni (L0545)
 Ne, ne bom več našteval (L0546)
 Vedno znova se pokaže, da je to utopija (L0547)
 Ljudje tudi mnogokrat optimistično pričakujejo, da bo popolnejši zakon prinesel tudi dobre rezultate (L0548)
 Imaš pri sebi fotoaparata (L0549)
 Na aeromitingu so prikazali nekaj zelo drznih letov (L0550)
 Hanca, meh za smeh (L0551)
 Dveh Peht pa že ne boš zmogel (L0552)
 Tank je vozil pred nami (L0553)
 Trd je kot kamen (L0554)
 Čisto se je moral sključiti, da je lahko sedel v avto (L0555)

Če ne boš ubogal, se boš moral cvreti v peklu (L0556)
 Naj vstane obtoženi Smlednik (L0557)
 To da bi bil vzrok za moje prijetje (L0558)
 Vdreti je želel skozi okno, a je ugotovil, da so vrata odprta (L0559)
 Zgledati je moral trezen, da ga ne bi odpeljali (L0560)

9.6 Corpus Lingua 3B

Da bi jim povedal novico, je dal sklicati vse člane posadke (L0561)
 Vinske trte ni hotel škropiti, zato je imel slab pridelek (L0562)
 Samo včlaniti se morate pa lahko kupujete knjige ceneje (L0563)
 Sljuda je izolator (L0564)
 Kakšen smrad pa je to (L0565)
 Vknjižiti je dal vse prihodke in odhodke (L0566)
 Kramp si zadeni na ramo in pojdi (L0567)
 Celo goro torb so mu pripeljali (L0568)
 Kam skakljaš (L0569)
 Oče je odšel po poljski poti (L0570)
 Zavila sta v ozko ulico (L0571)
 Desetletja vrti in presnema in preprodaja brez plačila ali z zelo majhnim plačilom vse, kar se vrteti da (L0572)
 Brezciljnež iz čiste lenobe (L0573)
 Ljudje se ne dajo (L0574)
 Če bom šel dovolj daleč, mi ne bo nihče sledil (L0575)
 Je hotel pokazati, da ima stil (L0576)
 Tudi njenega vprašanja mi ni treba izreči (L0577)
 Kaj pa vi (L0578)
 Vzdušje je dvignilo tudi izredno veliko občinstva, saj so poleg umetnikovih najbližjih znancev iz hiše prišli tudi mnogi drugi (L0579)
 Prej sem ga lomil s po tremi dekleti do pete ure zjutraj (L0580)
 Vendar pa čas obnove muzeja ne bo povsem mrtev, ampak se bo kljub zaprtosti stalne zbirke ka dogajalo (L0581)
 Zveza namerava še posebej uspešnim pridelovalcem konoplje podeljevati pokale (L0582)
 Ščegetal jo je, da je že povsod čutila ščemenje (L0583)
 Tako je bilo (L0584)
 Različne aplikacije zahtevajo različne stopnje vstopa v navidezni svet (L0585)
 Program je bil po številu namestitvenih disket in porabi prostora na disku skromen (L0586)
 Skenerji so bili svoje čase zgolj domena profesionalcev (L0587)
 Novo podjetje ima še več načrtov (L0588)
 Skvariti mu uspe še tako dobro zabavo (L0589)
 Vzmet je nenadoma izskočila iz ležišča in ni je več našel (L0590)
 Ne bodi tako vzvišen, saj vemo čigav si (L0591)
 Zmnožiti ni znal niti dveh števil (L0592)
 Zmračiti se mu je moralo, da se je sam odpravil na tako dolgo pot (L0593)
 Gora ni moč vzljubiti čez noč (L0594)
 Koliko gajb piva ste spili to noč (L0595)
 Z reševanjem teh enačb ne boste prišli daleč (L0596)
 Hlapec hodi h Golobovim v hram (L0597)
 Acetilena je zmanjkalo (L0598)
 Aha, si rečem, nekaj se bo napletlo, in to bo moja zasluga (L0599)
 Znamkica naj bi dosegla, da bi piratom odklenkalo (L0600)
 Človek se najprej seveda vpraša, čemu je bil prišel (L0601)
 Cesta je polna duhov (L0602)
 Me prav zanima, koga bo nadlegoval potem (L0603)
 On bi še kar govoril (L0604)
 Bil sem prijazen s tabo (L0605)
 Združeni narodi so objavili rezultate študije o plačah žensk (L0606)
 Zakaj ni bilo nikogar iz ljubljanskega baleta (L0607)
 Odstotki dajatev se dvigajo v okviru te lestvice (L0608)
 Na muzejskem dvorišču bodo tako pripravili otroško muzejsko delavnico (L0609)
 Praznujemo drugo obletnico rojstva samostojne slovenske države (L0610)
 Ščepec tobaka je vzel (L0611)

Težko breme težko nosi (L0612)
 V računalniški industriji že dalj časa velja nenapisano pravilo, da se v letu in pol vse podvoji (L0613)
 Z nekaterimi formati datotek program dela neposredno, druge pa lahko samo uvozi oziroma izvozi (L0614)
 Ploske skenerje rabimo v najrazličnejše namene (L0615)
 Kaj hitro smo ugotovili vse prednosti take rešitve, ki je nekoliko dražja od konkurenčne (L0616)
 Kakšno ločljivost pa sploh potrebujemo (L0617)
 Pri izdelavi ali spreminjanju strukture tabele določamo tudi indekse (L0618)
 Naenkrat lahko delamo le z eno bazo podatkov, vendar ni to nikakršna omejitev (L0619)
 Meso je bilo močno popoprano (L0620)
 Sosed je imel kozla in dve kozi (L0621)
 Krt se je skrtil na naš vrt (L0622)
 Največ denarja je prispevala občina, manjši del bo zbran v krajevni skupnosti (L0623)
 Organizirali bodo tudi smučarski tečaj (L0624)
 Hrovatova je na izjemno zahtevnem veleslalomu s sedmim mestom dosegla svoj najboljši rezultat v tej disciplini (L0625)
 Vsekakor nam mora biti zdaj že povsem jasno, kako politično težko, gospoarsko izredno drago, vojaško pa zelo tvegano je posegati v spor (L0626)
 Še bolj kot v kupoprodajnih odnosih povedano velja za druge odnose med podjetji in prizadetimi ljudi (L0627)
 Ne, v teh očeh ni privoščljive nežnosti (L0628)
 Priskrbel vam bom zdravila (L0629)
 Rad bi se šel oprhat (L0630)
 Najlažje je bežati, v prisposobi ali zares, najlažja možnost pa je najbrž tudi edina (L0631)
 Majhna država med večjimi, tujejezičnimi sosedi je kot nalašč za posredovanje kulturnih izdelkov (L0632)
 Potem spet molčimo (L0633)
 Že spet afne gunca (L0634)
 Preden stopiš predenj, se skrbno napravi (L0635)
 Tale asfalt je precej dotrajan (L0636)
 Naštevaj koliko tajg poznaš (L0637)
 Koliko krivd boste še spregledali (L0638)
 Seme je vzkliklo čez noč (L0639)
 Njegova vljudnost mi že preseda (L0640)
 Nenehno ga moraš gnjaviti, da kaj naredi (L0641)
 Vkliniti se pa ne dam (L0642)
 Režiser je moral vplesti še tretjega junaka, da je ohranil napetost zgodbe (L0643)
 V starih časih so majhne otroke povili v štruco, da so jih lahko postavili v kot (L0644)
 Vse, kar smo povedali doslej, se je nanašalo na skeniranje fotografij (L0645)
 Kalibracijo si najlažje predstavljamo pri sivinskih slikah (L0646)
 V omrežnem okolju nam ni treba skrbeti za osveževanje in zaklepanje zapisov (L0647)
 Svojo močno roko je položil na mojo (L0648)
 Jo, jojmene, jojmene in jojmenesta so medmeti (L0649)
 Nikoli ni stopil na pot (L0650)
 Vodo iz kanalizacije bodo pozneje speljali tudi skozi čistilno napravo (L0651)
 Nekaj mesecev pred iztekom mandata sedanje bi morala skupščina izvoliti celotno novo vlado (L0652)
 Najboljši slovenski tekači na smučeh, med njimi tudi nekaj biatloncev, se je pomerilo na prvi pregledni tekmi sezone (L0653)
 Vsake toliko nas je med hrumenjem motorja našlo čričkanje in žvižganje (L0654)
 Podoba, imidž podjetja se dela vsak dan v odnosih med podjetji in vpletenimi javnostmi (L0655)
 Celo lovec gleda odstreljeni plen s kančkom hvaležnosti, ponosa in razgretosti (L0656)
 Brez skrbi, vse se bo uredilo (L0657)
 Vidim, da vam je zmanjkalo cigaret (L0658)
 Ne, ne da se govoriti (L0659)
 Tonski studii so se razmnožili (L0660)
 Brezupno vesel sem, ne vem točno zakaj (L0661)
 Oba agenta sta prisluškovala pogovoru (L0662)
 Vse zaman, za manj ko sto ga ne dam (L0663)
 Deklič se preveč čička (L0664)
 Kot blisk je brzel mimo vseh vratc (L0665)
 Koliko Mick pa misliš da pride (L0666)
 Njega pa res ni težko vzdražiti (L0667)
 Spljuvati moraš vse kar imaš v ustih (L0668)

Zmleti moraš žito, da lahko spečeš kruh (L0669)
Pljuniti na tla je grdo (L0670)
Vzlic hudemu neurju so se odpravili na morje (L0671)
Hkrati ne morem poslušati obeh (L0672)

9.7 Corpus Lingua 4A

Ulovi sedem pišk pa dobiš petelina (L0673)
Spremljaj takt pa ne bo težav (L0674)
Prunč, prinesite še dva piva (L0675)
Takšnega kot je Zaverl pa ne najdeš povsodi (L0676)
Srh me spreleti, če samo pomislim na to (L0677)
V Sloveniji ni veliko koksarn (L0678)
Kot otrok sem imel mumps (L0679)
Leopardka ga je nemo opazovala s skale (L0680)
Dvanajst je odbila (L0681)
Njprej rahlo navzdol, potem strmo navkreber (L0682)
Ni eden od mnogih, je edini, je prvi (L0683)
Cele ure lahko postopam ali posedam in mi ni dolgčas (L0684)
Zmočilo se je tudi sedlo Letečega goloba (L0685)
Dostojanstveno se drži na smeh in zato ima lep obraz (L0686)
Ob vseh ograjah so namečkani potniki (L0687)
V petek bo večinoma sončno, le v gorskem svetu bodo popoldne posamezne nevihte (L0688)
Odpulula bo že jutri opoldne (L0689)
Med dokumenti, ki jih voznik včasih potrebuje na tujem, sta tudi mednarodn o vozniško dovoljenje in dovoljenje za vožnjo tujega vozila v tujini (L0690)
Najmanj sedem ljudi je izgubilo v gozdu (L0691)
Njegova književnost se je v večji meri uveljavila šele po drugi svetovni vojni (L0692)
Ker ni imel daru govora, se ni mogel z ničemer opavičiti (L0693)
Vdoblili so in vsekali znamenje v drevo (L0694)
Bilke iz skalnatih polj so ovenele (L0695)
Stric Silvo nad cesto deva nekaj pod zobe (L0696)
Srno prežene strah pred gromom (L0697)
Od izdajanja elektronskih knjig nimajo koristi samo založniki (L0698)
Resnici na ljubo gre velikokrat za slabe kompilacije (L0699)
Sistem vključuje vmesnik, ki poskrbi za to, da se računalnik prižge takoj, ko pozvoni telefon (L0700)
Pasovna širina zadostuje tudi za najbolj sanjske aplikacije (L0701)
Mestni park so razdejali vandali (L0702)
Gost je kralj le, če ima denar (L0703)
Harf v orkestru ni bilo moč slišati (L0704)
Koliko želv ste rešili to sezono (L0705)
Loščil je emajl, a nič ni pomagalo (L0706)
Valovi so bili previsoki in čoln se je prevrnil (L0707)
Brez Lenčk se ne da (L0708)
Poberi še tistih nekaj trsk (L0709)
Med njimi je bilo tudi osem kozmonavtk (L0710)
Skozi slečeno grmičje se vzpnem nanj in zadihan sedem na gobast kamen (L0711)
Obzirnost - do koga (L0712)
Povsod sem nekje na sredini reke, ki tiho teče od doma domov (L0713)
Petje in igranje privabita še več ljudi (L0714)
Vreščanje prihaja izpod debeločrtastih platnenih strešic, ki čepijo na pesku in se naslanjajo na zid (L0715)
Obzorje je vse bolj polno ladij vseh velikosti in stopenj zarjavelosti (L0716)
Tunizijski vojaki so na napad odgovorili in streljanja je bilo konec v petih minutah (L0717)
Ljudje se razlikujemo po potrebi po redu (L0718)
Običajno je v ustih veliko različnih bakterij, virusov in glivic (L0719)
O zdravstvenem stanju bolnice prihajajo nasprotujoče si vesti (L0720)
Okoli Ajdovščine še vedno piha močna burja, zato morajo s ceste vsi tovornjaki, ki so lažji od osem ton (L0721)
Njegov oče je tako hotel (L0722)
Vstopite, v veži je hladno (L0723)
Zdaj poglej v majsko jutro (L0724)
Do koč šaš sili (L0725)

Toper je res dober (L0726)
 Evropa je po konkurenčnosti za ZDA in Japonsko (L0727)
 Prepričanje, da je treba embalažo karseda skrčiti, zagovarjajo tudi prodajalci (L0728)
 Prvi projekt bodo vrata za paralelni priključek (L0729)
 Zakaj sploh potrebujemo navidezno resničnost (L0730)
 Razlike so bile še bolj opazne kot pri sivem klinu (L0731)
 Program namestimo s sedmih disket (L0732)
 Med pregledom lahko podatke tudi spreminjamo (L0733)
 Stolp ima šest prostorov za diske (L0734)
 Marca 1994 so razpisali natečaj za predloge raziskovalnih akcij v multimedijem založništvu (L0735)
 Prišla je k žetvi (L0736)
 Mlad oče nosi poln koš kož (L0737)
 Ena bo igrala, druga bo ušla (L0738)
 Naša Anka večkrat manjka (L0739)
 Drhal je vdrla v hotel (L0740)
 Komaj je napolnil pol tiskovne pole (L0741)
 V političnih krizah za poštenega človeka ni najtežje opravljati svojo dolžnost, ampak razumeti v čem obstaja (L0742)
 Kot sadje po vsem svetu, tudi tropsko sadje čeprav naprodaj v Evropi, ni zgolj snobovska modna muha, marveč tudi učinkovito osvežilo (L0743)
 Kljub velikemu tehnološkemu napredku človek še vedno ni kos mrazu (L0744)
 Šola jo je morala stati, kolikor se da malo (L0745)
 Omeniti moramo, da je streljaj od hiše akumulacijsko jezero, kjer bodo lahko uživali ribiči (L0746)
 Pohajati po ulicah je vselej nekaj novega (L0747)
 Smeji se in ne pokaže truda, da bi kaj zvedel od mene (L0748)
 Čutim, da bi še moral storiti nekaj (L0749)
 Z velikim kamnom v grlu in majhnim zavitkom v roki odidem proti izhodu (L0750)
 Paralel je še več, končajmo z naslednjo: vsi trije so povezani z Notranjsko (L0751)
 Ne morem vpeti strune, niti vdeti zatiča (L0752)
 Odnehal ni niti, ko je bil kos že popolnoma scvrt (L0753)
 Koliko majčk ste prodali lani (L0754)
 Takole molst se pa ne dam več (L0755)
 Po prešanju je ostal samo krš (L0756)
 Premalo fajf imate, da bi lahko vsi kadili (L0757)
 Spremenljivka iks naj zavzame samo tri vrednosti (L0758)
 Tole morate napraviti iz samih jelš (L0759)
 Na pismo ni potrebno nalepiti znamk (L0760)
 Jure Zdovc je eden naših boljših košarkarjev (L0761)
 Manjka mu še malo dodelave za lažje in lepše uporabniško delo (L0762)
 Trg se jim zdi stabilen (L0763)
 Ker so cene diskov približno enake, kakovostne razlike pa precejšnje, so kupci pogosto v zadregi kaj izbrati (L0764)
 Z novim podnožjem bi dosegli asimetrično povezavo (L0765)
 Sladka svatka je po svatbi padla pod pot (L0766)
 Zaplavaj k bregu (L0767)
 Pomisli, predenj stopiš predenj (L0768)
 Val je odnesel brv (L0769)
 Soobtoženka je globoko sopla (L0770)
 Redko uporabljamo kreme ali olja za roke (L0771)
 Kopalci so na vodi odkrili nezvestnega dečka, ki so ga v bolnici rešili (L0772)
 Snemanja večinoma potekajo v naravi in prinašajo tudi nove, prej nepredvidene situacije (L0773)
 Postavila sva zid vsak pred svoje izkušnje (L0774)
 Ob hiši sta tudi dve novi teniški igrišči, saj si turistične kmetije brez možnosti rekreacije ni mogoče predstaviti (L0775)
 Tu ni več kaj odpirati ust, lahko se samo ponavljam (L0776)
 Ustaviva se in si dava roki (L0777)
 Ni ga, čeprav se mi zdi, da je kar naprej z mano (L0778)
 Ob vsakem krilu nihajnih vrat stoji kot slika nemogoče lepa lepotica (L0779)
 Je zato Krpan Notranjec (L0780)
 Če se ji bo zljubilo, bo šla (L0781)
 Žbice mu zelo lepo dišijo (L0782)
 Samo aktivistk ne vprašajte (L0783)

Ob zori je prišel v Gdansk (L0784)
Matevž mu je nemo ponudil roko (L0785)
Olga Rems je napovedala novo oddajo (L0786)
Trs se vam bo posušil (L0787)
Himn se nismo učili na pamet (L0788)
Grušč se je splazil po strmini (L0789)
Prinesel je veliko risb, a mu nobena ni bila povšeči (L0790)

9.8 Corpus Lingua 4B

Koliko tipk ima pisalni stroj (L0791)
Verjameš, da so gradili babilonski stolp (L0792)
Mulc mu je pobegnil pred nosom (L0793)
Takah ujme pomnijo niti najstarejši ljudje (L0794)
Zbrali so se ljudje iz različnih branž (L0795)
Nenadoma mu je polh skočil za vrat (L0796)
Črv je prilezel iz luknje (L0797)
Njegov instinkt nas je gnal naprej (L0798)
Samo Uršk še ni bilo (L0799)
Kaj je dnjača (L0800)
Pogled se sproščujoče pase po gozdnih zaplatah in košeninah Mrzlega vrha (L0801)
Ali sodi ta preblisk v napleteno zgodnico (L0802)
To opazim spotoma, ne da bi pomislil (L0803)
Nikomur ne privoščim tega, česar ne poznam (L0804)
Kje imajo shranjeno pritožno knjigo (L0805)
Danski poštarji bodo posegli po tradicionalnih metodah pridobivanja prijateljev med hišnimi čuvaji (L0806)
Zakaj si tako razposajen (L0807)
Ugotovili so, da so ljudje zboleli potem, ko so dobili transfuzijo krvi iz ene od omenjenih bank (L0808)
Kadar postane neko dogajanje javno, ga je potrebno v javnosti tudi odkrito in v celoti predstaviti in skleniti (L0809)
Dva zavoja sem v tem položaju še speljala, zadnjega pa ne (L0810)
Med štirinajstimi dekleti je bila na zahtevni progi po pričakovanjih najhitrejša Nataša Lačen (L0811)
Pride v četrtek (L0812)
Manjši je boljši (L0813)
Sin noče izzvati oddelka (L0814)
Zjutraj pojdemo brez jeze (L0815)
Železni prijem recesije v računalniški industriji počasi popušča (L0816)
Vsi našti diskimajmo enak uporabniški vmesnik izdelan s hipertekstom (L0817)
Vsak vzorec lahko takoj premaknemo za oktavovavzgor ali navzdol (L0818)
Cilj je bil pokazati eno od možnih aplikacij navidezne resničnosti (L0819)
Z manjšo zamudo so začeli izdelovati tudi ročne skenerje (L0820)
Kupil si je več pečk, samo da ga ne bi zeblo (L0821)
Noben lišp mu ni več pomagal (L0822)
Dreč ga bomo klicali (L0823)
V mestu je polno malih prodajaln (L0824)
Koliko rajž že imaš za seboj (L0825)
Vrteli so moj najljubši film (L0826)
Zdolbst je moral veliko luknjo, preden je lahko sedel vanjo (L0827)
Francka mu je prva ponudila roko (L0828)
Koliko Julčk je med vami (L0829)
Rože so vcvetele navkljub vremenu (L0830)
Ne vem, ali bom znal, ali bom zmogel, odzvanja v meni (L0831)
Doklej še (L0832)
Zaboga, to bi lahko počel ves dan (L0833)
Vprašajte njihove sosede (L0834)
Prevajalec je ves cunjast od napora (L0835)
Odšel je na bojišče kot navaden vojak, ne kot pisatelj in intelektuallec (L0836)
Zraven boš, ko ji jo bom dal, tako si je ne bo upala odkloniti (L0837)
Pogoji za medenje so izredni, potrebno je le toplejše in bolj vlažno vreme brez jugozahodnika (L0838)
Sonce prodre tudi čez prosojno ruto, ki poskrbi za rahlo porjavlost brez opeklin (L0839)
V arabskem svetu resda cenijo denar, toda z bankovci kljub temu ne ravnaajo preveč spoštljivo (L0840)
Ko sediš v zasedi, si dobro oglej popotnika, ki prihaja (L0841)

Vsak vzame zmeraj vse (L0842)
 Duševna bol bolj boli (L0843)
 Ohol lenuh ob pitju sede na suh hlod (L0844)
 Tre precej trde orehe (L0845)
 Čim več ukazov mikroprocesor lahko izvrši hkrati, tem bolj kompleksen je vgrajeni tolmač (L0846)
 Programi se večinoma ne ponavljajo (L0847)
 Še vedno pa nameravajo podpirati obe verziji (L0848)
 Kaj se je spremenilo in kako (L0849)
 Omenjena rešitev pa terja tudi svoj davek (L0850)
 Najprej so se pojavile preglednice (L0851)
 Pred leti so se pojavili grafični uporabniški vmesniki (L0852)
 Zadevo lahko primerjamo z razcvetom žepne knjige po drugi svetovni vojni (L0853)
 Potencialna moč Evrope je v bogatih vsebinah informacij, ki so se nakopičile zaradi preobilja pisnih in drugih gradiv (L0854)
 Odpravlja se k Dravi (L0855)
 Slok telovadec ima dober slog (L0856)
 Marec že prihaja (L0857)
 Vsekakor, v sobi je vsak varen (L0858)
 Polde je govoril, da pojde ob visoki vodi (L0859)
 Ob polnoči je postal pol človek pol konj (L0860)
 Konstruktorji za člane se kličejo v istem vrstnem redu, kot so zapisani v deklaraciji razreda (L0861)
 Sintagmatika glasov obravnava pojave, ki se pri glasovih pojavljajo zaradi soseščine drugih glasov (L0862)
 Pijača ima posladek okus in v poletni vročini kar dobro dene (L0863)
 Zakaj bi nekaj počel drugače, kot sem doslej (L0864)
 V prilogi, ki bo v sredini časopisa, bomo lahko izvedeli več o raznih informacijah, pomembnih za študente (L0865)
 Ulice so polprazne in to človeka moti (L0866)
 Galebi sedajo okrog njih, vendar nikoli preblizu (L0867)
 Vse se blešči od slonokoščene beline zob (L0868)
 Nekateri kričijo, nekateri se zabavajo (L0869)
 Nekaj malega bom pobrskal po tej ropotarnici, ki še čaka zgodovinarja - redarja (L0870)
 Če je tako, pa nasvidenje (L0871)
 Izdolbst je moral še dve luknji (L0872)
 Živ krst ga ni našel (L0873)
 To je bil že drugi infarkt (L0874)
 Samo puls vam še izmerim (L0875)
 Školjč naj nam razloži (L0876)
 Krč ga je zgrabil ko je bil še daleč od obale (L0877)
 Na voljo vam je deset različnih nians (L0878)
 Gospod Kant, kaj vam smem ponuditi (L0879)
 V temi nisi mogel videti drobcenih telesc (L0880)
 Na trgu je kar precej programov v različnih ovitkih (L0881)
 Pri delu so nam v pomoč ikone (L0882)
 V Statistiki najdemo vrsto podatkov in primerjav o našem dokumentu na dveh zaslonih (L0883)
 Izvedenci za snovanje iger se bodo usmerili k oblikovanju izobraževalnih in referenčnih izdelkov (L0884)
 K zidu se obrača (L0885)
 Dekle košatih las odhaja v sosedov laz (L0886)
 Proč že odžira (L0887)
 Ubil je muho in jo vrgel v smetnjak (L0888)
 Za pravdo je prav dobro pripravljen (L0889)
 Po pol osmih je še vozil (L0890)
 V drugem nastopu je Hrovatova smučala v slogu velikih šampionk (L0891)
 Poleti najlažje poskrbimo za svoje telo z gibanjem, zato na morju čimveč kolesarite, plavajte in tecite (L0892)
 Pisateljica romanov piše o svojih sanjarjenjih, zgodbe iz njenih romanov pa se vseskozi prepletajo z njenimi lastnimi doživljaji (L0893)
 Konrad je bil neodgovoren otročaj, bister, a bolešno narcisoiden (L0894)
 V Prekmurju je turizem na vasi, v primerjavi z drugimi deli Slovenije, še v povojih (L0895)
 Marsikdaj sem takšen, da bi me bilo vredno skriti - vendar ne nocoj (L0896)
 Odkritosrčen nasmeh to pove (L0897)
 Kaj je narobe s temi ljudmi, najmanjšega znaka življenja ne dajo od sebe (L0898)
 Celo papir prenese vse (L0899)

Dva med njimi sta meščana, tretji pa podeželan - vsi pa revnih staršev (L0900)
 Gvaš je njegova najljubša tehnika (L0901)
 V tem kolu je bilo sedem dvanajstk (L0902)
 Do včeraj je bilo že sedem izvržb (L0903)
 Na barju je veliko najdb (L0904)
 Žekš in jaz nikoli nisva stopila noter (L0905)
 Odstraniti so mu morali precej limf (L0906)
 Mama, stric Dolf je prišel (L0907)
 Če porečeš kokodajs še ne pomeni, da boš tudi znesel jajce (L0908)
 Cesarstvo je propadlo zaradi spletk (L0909)
 Čemu ti bo toliko cevki (L0910)

9.9 Corpus Lingua Passage

V svetu je danes v uporabi več kot 120 milijonov osebnih računalnikov (P0911)
 Medtem ko večina sveta šele vstopa v dobo računalnika, je ta v Evropi, na Japonskem in v Ameriki že postal nepogrešljiv vsakdanji pripomoček (P0912)
 Strojna oprema postaja vse hitrejša (P0913)
 Programska oprema omogoča uporabo računalnika na skoraj vseh področjih (P0914)
 Razvoj tehnologije in inovacije vsako leto odpirajo vedno boljše in cenejše možnosti (P0915)
 Vse to omogoča uporabo računalnika tudi v najmanjših podjetjih (P0916)
 Še pred petimi leti je nabava računalnikov, tiskalnikov, optičnih čitalnikov veljala za veliko naložbo, danes pa je to za podjetje majhen zalogaj (P0917)
 Obdelava podatkov z računalnikom, uporaba preglednic, pošiljanje sporočil po faxu in podobno, so danes povsem vsakdanje računalniške funkcije (P0918)
 Poglejmo si nekaj možnosti uporabe računalnika v Ameriki, kjer je ta že postal integralni del poslovnega življenja (P0919)
 Večina načrtovanja in risanja se zdaj opravlja z računalnikom (P0920)
 Inženirji lahko z računalnikom ocenijo kritične točke pri mostovih (P0921)
 Pri samem načrtovanju pa vam tak program tudi samodejno izvaja izračune za statično trdnost konstrukcij (P0922)
 Če gradite novo hišo, igrišče za golf, nogometni stadion ali bazen, je prvi korak uporaba računalnika (P0923)
 Ta vam bo na zaslonu pričaral vse možne kombinacije, barve, vzorce, še preden je zasajena prva lopata (P0924)
 Tudi po končani gradnji vam računalnik lahko pomaga pri načrtovanju notranje opreme stavbe (P0925)
 Na zaslonu vam prikaže stanovanje opremljeno na najrazličnejše načine, z različnih zornih kotov in v vseh barvnih odtenkih (P0926)
 Velike trgovske verige v Ameriki že dlje časa uporabljajo računalnike pri prodaji (P0927)
 V avtomobilskem oddelku vam z ukazom na dotik omogočijo, da najdete ravno tisti del za vaš avto, ki ga potrebujete (P0928)
 Preprosto pritisnete na gumb z možnimi izbirami za vrsto avta, znamko, model, letnik (P0929)
 V malo več kot minuti se na zaslonu izpiše, kateri tip oljnega filtra potrebujete ali kakšne svečke so primerne za vaš avto (P0930)
 V prodajalni videokaset vam osebni računalnik hitro pomaga najti željeno kaseto (P0931)
 Iščete lahko po naslovu, zvrsti filma, ali celo po imenu vašega priljubljenega igralca ali igralk (P0932)
 Pri izbiri pohištva vam računalnik tudi svetuje kakšna omara ali miza bi vam najbolj ustrezala glede na vašo trenutno opremljenost stanovanja (P0933)
 Večina smučarskih središč v Koloradu ima smučišča urejena po sistemu črtne kode (P0934)
 Ta je odtisnjena tudi na karti, to pa pomeni, da brez nje ne morete na sedežnico ali vlečnico (P0935)
 Tudi goljufije s ponarejenimi kartami niso več možne (P0936)
 Smučarski delavci imajo tako vsak dan točen pregled nad prodanimi kartami (P0937)
 Natančno vedo, koliko je eno ali več dnevnik, koliko jih je s popustom, na katerem smučišču se smučarji največ zadržujejo (P0938)
 Vedo tudi koliko časa smučarji porabijo za kosilo, kdaj se zanj odločajo in koliko časa na dan aktivno smučajo (P0939)
 Takšne informacije vodstvu smučarskega centra pomagajo prepoznati profile smučarskih potrošnikov (P0940)
 Največje verige hitre prehrane uporabljajo računalnike v podporo svoji razvojni strategiji in za dostavo hrane na dom (P0941)
 Računalnik zabeleži vsako naročilo pizze od doma, in ko kupec pokliče naslednjik, je njegov naslov že na zaslonu (P0942)

Če kupec dlje časa ni naročil pizze, mu pošljejo prijazno vabilo, naj spet naroči pri njih (P0943)

Osebu tako ni treba tratiti časa z zapisovanjem podatkov (P0944)

Takšna banka podatkov je koristna tudi pri promocijskih dejavnostih podjetja (P0945)

Pri odprtju novega lokala ima podjetje takoj na voljo naslove vseh svojih naročnikov, da jih o tem obvesti (P0946)

Celo natakari že začenjajo nositi žepne računalnike, s katerim lahko v hipu prenesejo naročilo od gosta v kuhinjo (P0947)

Zlasti je tak računalnik uporaben kadar gost naroči kakšen zahteven koktajl ali jed, ki terja zamudno pripravo (P0948)

Tudi ko kakšne hrane v kuhinji ne morejo pripraviti, je natakari o tem sproti obveščen (P0949)

Tako lahko uspešno odpravimo nepotrebno tekanje natakarijev od kuhinje in nazaj, ki sicer zahteva kar 60 odstotkov natakarijevega časa (P0950)

Ko je jed pripravljena, dobi natakari obvestilo na svojem računalniku in jo lahko takoj postreže gostu (P0951)

Računalnik tako lahko zelo pohitri gostinske storitve, pa tudi napitnine so večje, saj natakari lahko več časa posveti gostu (P0952)

Gasilci in policija veliko in uspešno uporabljajo geografski informacijski sistem (P0953)

Grafični zemljevidi omogočajo izbiro najkrajše poti do kraja nesreče (P0954)

Postrežejo tudi s podatki o prometnih zamaških, konstrukciji cestišča in cestnih odsekih s težkimi tovornjaki (P0955)

Vodje reševanja lahko na monitorju spremljajo tudi potek akcije in usklajujejo delo reševalnih ekip (P0956)

Obdelava statističnih podatkov je dobrodošla tudi pri pošiljanju patrolnih vozil na pričakovana mesta nesreč ali kriminalnih dejanj (P0957)

Veliko podjetij porabi veliko časa in denarja za pripravo, izdelavo in izdajo prodajnih katalogov, prospektov in drugega reklamnega gradiva (P0958)

Zdaj si kupec lahko ogleda vse izdelke na zaslonu prek računalniške mreže in modemskih povezav, v barvah ter ob zvočni spremljavi (P0959)

Postopek naročanja se je zelo skrajšal (P0960)

Nič več ni treba iskati cenikov in jih po faxu ali po pošti pošiljati potencialnemu kupcu ter čakati na naročilo (P0961)

Vse to lahko opravite z računalnikom v pičlem dnevu (P0962)

Elektronska pošta vam omogoča, da kar najhitreje dobite informacije o zadnjih cenah, popustih ali promocijskih dejavnostih (P0963)

Zelo razširjeni so računalniški projektorji, ki prodajalcem omogočajo predstavitve brez zamudne uporabe diapozitivov (P0964)

Če povežete s svojim prenosnim računalnikom običajni projektor ali grafoskop, se bo vsebina vašega monitorja pojavila tudi na velikem zaslonu (P0965)

Za kakršnekoli popravke zadošča že pritisk na tipko (P0966)

Morda delate pri projektu, ki zahteva veliko povratnih informacij, obdelav in revizij (P0967)

Osební računalnik vam dovoljuje, da se prek modemskih povezav udeležite konference, kjer si izmenjujete informacije s skupinami z različnih koncev sveta (P0968)

Namesto da bi vam poslali željeno informacijo po faxu, se ta izpiše na vašem zaslonu (P0969)

Vanjo lahko vnesete spremembe ali dopolnila (P0970)

Popravljená informacija je nato na voljo tudi drugim uporabnikom (P0971)

Za 80 dolarjev vam špedicijski računalnik ponuja stalne informacije o prevoznih tarifah (P0972)

Za vsako linijo v ZDA imate na voljo informacije o cenah prevozov pri konkurenčnih prevoznikih po morju, cesti, zraku ali železnici (P0973)

Programi vam tudi svetujejo najustreznejšo vrsto prevoza, pot in čas glede na tip tovora in vaše zahteve (P0974)

Nekatere od največjih svetovnih založniških hiš zdaj ponujajo tudi posebne banke podatkov (P0975)

Te banke vsebujejo sezname, kjer so navedeni podatki o vseh delih založbe (P0976)

Poleg naslovov in avtorjev lahko dobite tudi kratke povzetke posameznih del, poglavij ali člankov (P0977)

Preko računalniških mrež lahko pošlejte založniku zahteve po določenem gradivu (P0978)

Založnik nato pripravi željeno gradivo in ga v računalniški obliki pošlje ravno tako po mreži (P0979)

Celoten postopek tako traja le nekaj dni namesto več tednov ali mesecev (P0980)

Za potrošnika so možnosti uporabe računalnika zares neverjetne (P0981)

Na preizkušnji je tudi vrsta futurističnih in fantastičnih uporab, vendar je usodnega pomena tako imenovana informacijska avtocesta (P0982)

Ta bo številnim podjetjem omogočala dostop do pomembnih informacij (P0983)

Javna in izobraževalna mreža INTERNET že elektronsko povezuje okrog dvajset milijonov ljudi po vsem svetu (P0984)

Podobna mreža v Evropi EDI postaja priljubljena tudi v Ameriki (P0985)

Preko te mreže si že okrog štirideset tisoč podjetij izmenjuje račune elektronsko (P0986)
 V nastajanju je tudi program CommerceNet, ki ga podpirajo podjetja kot so Apple, Intel in HP (P0987)
 Ta bo povezoval kupce, prodajalce, proizvajalce in dobavitelje visoke tehnologije (P0988)
 S pomočjo tega programa bo moč preprosto in hitro poiskati najboljše cene in najugodnejše ponudbe (P0989)
 Ravno tako bo z njim možno spremljati povpraševanje o izdelkih, podpisati pogodbo in pospešiti plačilo (P0990)
 Zelo malo ljudi, ki danes pri svojem delu uporabljajo računalnik, je z njim tudi odraslo (P0991)
 Naslednje generacije pa bodo izkušnje z računalniki na delovno mesto prinesle že iz otroških let (P0992)
 Več držav v ZDA načrtuje, da bi v naslednjih šestih letih računalnike imel vsak razred osnovne šole (P0993)
 Kompaktne plošče z raznimi bazami podatkov omogočajo otrokom hiter dostop do ogromnega števila informacij (P0994)
 Generacije, ki bodo zrasle s toliko podatki na dosegu svojih prstov, bodo tako tudi pri svojem poklicnem delu zahtevale podoben pristop (P0995)
 S tem bodo pomembno spremenile način dela v prihodnosti (P0996)
 Ocenjujejo, da ima tretjina ameriških gospodinjstev doma osebni računalnik (P0997)
 Poznavalci razmer napovedujejo, da bo do leta dva tisoč vsako ameriško gospodinjstvo imelo že dve cela pet osebnega računalnika (P0998)
 Razširjenost računalnika pa še povečuje napredek v tehnologiji, ki povečuje njegove zmogljivosti (P0999)
 Tudi računalnikovo razumevanje pisave, slike ali govora se počasi seli iz filmov v vsakdanjost (P1000)
 Lahko si že kar predstavljamo, kakšen bo svet poslovnega življenja čez nekaj desetletij (P1001)

10. Set of used words, numbers and set of digits for corpus WORDS

10.1 Corpus WRD

končaj (W1533)
 počasneje (W1534)
 skoči (W1535)
 prekini (W1536)
 Amsterdam (W1537)
 Bukarešta (W1538)
 Luksemburg (W1539)
 Rim (W1540)
 levo (W1543)
 zgrabi (W1544)
 preskoči (W1545)
 Maribor (W1546)
 Andora (W1547)
 Dublin (W1548)
 Madrid (W1549)
 Sofija (W1550)
 start (W1553)
 nazaj (W1554)
 vključi (W1555)
 napaka (W1556)
 Velenje (W1557)
 Bruselj (W1558)
 Lizbona (W1559)
 Praga (W1560)
 začni (W1563)
 hitreje (W1564)
 izključi (W1565)
 ponovi (W1566)
 Lendava (W1567)
 Budimpešta (W1568)
 London (W1569)
 Reykjavik (W1570)
 Stockholm (W1571)
 Monako (W1572)
 Dunaj (W1573)
 Atene (W1574)

Ljubljana (W1575)
sporoči (W1576)
spusti (W1577)
desno (W1578)
da (W1579)
Tirana (W1581)
Moskva (W1582)
Helsinki (W1583)
Bern (W1584)
Koper (W1585)
preglej (W1586)
odpri (W1587)
gor (W1588)
ne (W1589)
Vaduz (W1591)
Oslo (W1592)
Istambul (W1593)
Bonn (W1594)
Celje (W1595)
iztiskaj (W1596)
zapri (W1597)
dol (W1598)
stop (W1599)
Varšava (W1601)
Pariz (W1602)
Kopenhagen (W1603)
Bratislava (W1604)
Krško (W1605)
izpiši (W1606)
opusti (W1607)
naprej (W1608)
ustavi (W1609)

10.2 Corpus ABC

Š (A1611)
R (A1612)
Č (A1613)
V (A1614)
K (A1615)
Z (A1616)
J (A1617)
H (A1618)
T (A1619)
X (A1620)
B (A1621)
N (A1622)
E (A1623)
P (A1624)
Ž (A1625)
D (A1626)
L (A1627)
A (A1628)
M (A1629)
U (A1630)
S (A1631)
I (A1632)
F (A1633)
C (A1634)
O (A1635)
G (A1636)

10.3 Corpus SETS

dve ena dve sedem ena pet (S1637)
osem devet nič sedem sedem štiri (S1638)
tri nič šest osem pet ena (S1639)
sedem nič devet nič pet devet (S1640)
ena nič dve nič dve ena (S1641)
štiri štiri pet dve sedem ena (S1642)
štiri tri šest nič tri tri (S1643)
dve ena ena pet štiri osem (S1644)
dve nič štiri tri dve osem (S1645)
dve pet tri sedem ena devet (S1646)
sedem nič šest šest devet tri (S1647)
šest pet pet ena osem dve (S1648)
pet ena dve tri pet ena (S1649)
štiri ena pet dve štiri tri (S1650)
osem osem devet sedem ena pet (S1651)
tri štiri dve dve pet sedem (S1652)
nič osem dve šest devet devet (S1653)
pet ena nič osem sedem ena (S1654)
tri osem ena sedem dve šest (S1655)
nič nič ena šest tri devet (S1656)

10.4 Corpus NUMBERS

štiri (N1531)
osem (N1532)
pet (N1541)
devet (N1542)
dve (N1551)
šest (N1552)
tri (N1561)
sedem (N1562)
nič (N1580)
nula (N1590)
ena (N1600)
dva (N1610)