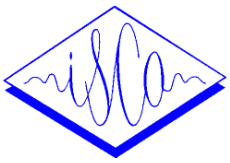

ISCA & IEEE Workshop on Spontaneous
Speech Processing and Recognition
Tokyo Institute of Technology, Tokyo, Japan
April 13-16, 2003

THE MAP TASK CORPUS OF SPOKEN RUSSIAN

Veronika Makarova[1], *Valery A. Petrushin*[2]

[1]Meikai University, National Institute of Advanced Industrial Science and Technology, Japan;
makarova.veronika@aist.go.jp
[2]Accenture Technology Labs, Accenture, Chicago, USA; petr@techlabs.accenture.com

ABSTRACT

This paper describes the purposes, structure and applications as well as the speakers, material, recording, digitizing, labeling and storage procedures of Map Task Corpus of spoken Russian. The database is comprised of recordings of 116 spontaneous unscripted task-oriented dialogues produced by 64 native speakers of Russian while performing the task of marking a route on a printed map. The task was performed only via verbal communication, other forms (such as eye contact or gestures) being excluded in experimental settings. The total duration of the recorded dialogues is 18 hours. The database is constructed as a material source for theoretical and applied linguistics, language teaching, psycholinguistics, communication, speech processing, recognition as well as for interdisciplinary applications.

1. INTRODUCTION

1.1. Large speech corpora, their use and challenges

The new focus in speech processing research on dialogue and communication systems has lead to an increased interest in various types of natural speech corpora, such as meetings [4]. Large corpora are an extremely rich source of data for phonetic research providing insights into the variability of segmental and suprasegmental units [6,8]. Furthermore, storage and processing of large computer corpora of spoken languages are required for extracting segmental prosodic and intonation parameters for speech synthesis, for training automatic speech processing systems, for conducting various forms of linguistic and psycholinguistic analysis and for applications in traditional and computer-assisted foreign language teaching. We therefore see a steady increase in the number of large natural speech corpora and in the scope of languages covered. Large corpora are time and effort-consuming, in particular, manual phonetic annotation becomes impossible, and has to be replaced by automatic transcription [6]. The nature of spontaneous speech sets a range of specific challenges to researchers working on large corpora. Spontaneous speech is hard to obtain in laboratory conditions, whereas field recordings may have faulty sound quality. A totally uncontrolled spontaneous speech can be incomparable across pairs of speakers who would choose different topics or deviate from the topics in an unpredictable way. The wish to obtain speech samples which would be studio-made, partly controlled, but yet spontaneous (or at least quasi-spontaneous) has lead to the creation of Task Corpora, whereby the subjects are assigned some kind of task (a game, discussion, video clip or picture description, etc.) requiring verbal input from the participants [4]. Examples of these are map task corpora.

1.2. Map task corpora

Map task corpora have been created for a few languages including different varieties of native and non-native English [1, 7], Japanese [9], Portugese [5], etc. They are based on the verbal interaction between two participants engaged in a task which requires one participant to mark a route on a map following the partner's instructions. The exact details vary across corpora, e.g., participants may have identical or somewhat different maps, there can be various conditions for lack/presence of eye contact, and for other features of experimental settings. The corpora are designed not only as sources for linguistic, applied linguistic and speech processing research, but also for multiple interdisciplinary purposes, such as investigations of problem solving strategies or studies of speech under various psychological, medical and physical conditions [2].

1.3. The advantages of map task corpora

While pertaining the spontaneity of participants' speech due to their concentration not on the verbal input per se, but on the performed task, such corpora allow researchers to overcome some quantitative and qualitative limitations of purely spontaneous dialogues [1]. In particular, the restriction of the topic enables

researchers to obtain speech samples compatible not only across speakers, but across corpora and languages. Inserting controlled phonetic features into the names of landmarks on the map makes it possible to study a particular phonetic phenomenon represented in a natural conversation.

2. THE CONSTRUCTION OF MAP TASK CORPUS OF SPOKEN RUSSIAN

2.1. Maps

Four quartets of maps were created for the database by replacing English landmark names on the maps employed in the HCRC and Japanese Map task corpora [1, 9] with equivalent Russian names. Russian names assigned to landmarks were selected so as to represent phonetic features relevant for the sound system of Russian and to match the actual picture images. Every quartet consisted of 4 pairs of maps, whereby one map in every pair had a route marked between landmarks (the Guide's map), and the other map did not have the route (the Traveler's map). The total number of maps was therefore 32 (2 maps x 4 sets x 4 quartets). The Guide's and Traveler's maps in every pair had a few small differences in the position and type of the landmarks.

2.2. Phonetic features

The phonetic features incorporated into the maps allow researchers to study a wide range of segmental, prosodic and intonation features in connected speech.

Segmental features

1) Vowel reduction and unstressed vowels (chastokol)
2) Assimilation (vhod v fort, prud s flamingo)
3) Phonetic manifestations of flexions (jestestvennoje ozero, rasshyrennyj put')
4) Vowel combinations or phonetic diphthongs (oazis, piramida faraona, babuiny). It should be noted that phonetic diphthongs are relatively rare in Russian which tends to avoid vowel combinations.
5) Palatalization in consonants not immediately preceding a front vowel (peshehodnyj mostik, dom bez dverej)
6) Alternative phonemic structures of words (kemping; razvalivshijsja skvorechnik, stolb-totem)

Suprasegmental features

7) The manifestation of stress and degrees of stress in compound words (vodosbor, zheleznodorozhnyj pereezd; polurazvalivshajasja usad'ba)
8) Accent assignment rules in phrases (mesto krusheniya sverhzvukovogo samol'ota, pljazh s bulyzhnikami, neskol'ko grjad gor)

Intonation features

9) Realizations of accents in words with different segmental structures and position of lexical stress (ozero – ozerco)
10) Realization of contrast (contrastive accents), i.e. when either the landmark names in the Guide's and Traveler's maps had a slight difference, or there were two landmarks with similar names on both maps (izgib pritoka – izgib potoka, laz v vorotah – prolom v vorotah, kop' s sol'ju – kop' s zolotom)
11) The database allows researchers to undertake studies of various intonation features (tone-unit size and structure, tones, accent assignments, etc).

2.3. Tasks

Two instruction sheets were prepared: one for the Guide and the other for the Traveler. Both instructions sheets were entitled "The rules of the game" and specified the roles of the respective participants. Both sheets said that the route is the only safe way to pass through a dangerous area and warned that the maps may have slight differences. The Guide was encouraged to explain the route with full detail, and the Traveler to ask anything he/she may have problems with.

2.4. Subjects

The database contains recordings of 116 dialogues by 64 native speakers of Russian: 12 male and 52 female. All of them were students of the Linguistics Faculty, St. Petersburg State University.

N Of Speakers	Speakers' age in full years									
	16	17	18	19	20	21	22	24	25	28
Female	0	12	21	9	3	1	0	2	0	1
Male	1	1	4	2	1	0	2	0	1	0
Total	**1**	**13**	**26**	**12**	**4**	**1**	**2**	**2**	**1**	**1**

Table 1: Speakers' distribution by age and gender

Fifty-four speakers have lived in St. Petersburg since their birth or early childhood. Ten speakers have lived in St. Petersburg for at least three years before the time of the recording, but come from other regions. Two of them come from Tver' (central Russia), and one from each of the following locations: Blagoveschensk (Amur region of Siberia), Chelyabinsk (the Urals region), Cherepovets (Northern Russia), Dnepropetrovsk (Ukraine), Klaipeda (Lithuania), Murmansk (Northern Russia), Ulyanovsk (Midland Russia near the Volga river), and Vorkuta (Northern Siberia). The average age of speakers is 18.7 years (s.d. = 2.15), with the range from 16 to 28 years. The speakers' distribution by age and gender is represented in Table 1.

2.5. Recording procedure

Environment

All recordings were performed in a sound-proofed studio of the Department of Phonetics, St. Petersburg State University, St. Petersburg, Russia. The subjects were seated on the opposite sides of two adjacently placed desks with a cardboard screen inserted between the desks to prevent the eye contact and gesture exchange. The subjects communicated via Sennheiser headsets, their voices were simultaneously recorded and transmitted to the partner over earphones. The subjects' behavior was monitored from the outside through the studio window to ensure that they followed the rules of the game. The grouping of the subjects was not controlled by gender and by degree of familiarity (i.e. whether they met before or not).

Equipment

All the data were recorded on a portable Digital Audio Tape-recorder Sony TCD-D8 at 48 kHz sampling rate via Sennheiser headphone sets in a sound proof recording studio.

2.6. Digitizing and storing the dialogues

The recordings were converted into stereo Windows PCM format at 32 kHz sampling frequency and 16 bits resolution. Each speaker possesses his/her own channel in stereo recordings. Each dialogue is stored in a separate file with the name that identifies the subjects and the task (maps) used. Once fully annotated, the Corpus will be prepared for release on CD-ROM disks.

2.7. Labeling of the corpus

Since each speaker has a separate channel, we can apply simple signal processing techniques for speaker tracking and calculating number and durations of turns. These data serve as initial speaker identification layer for dialogue transcription with Transcriber [3]. This manual tool is employed for classifying each turn according to its function (instruction, clarification, confirmation, etc.). The use of separate speaker channels also simplifies applying automatic speech recognition engines and prosody labeling techniques. Currently, we are working on speaker and turn identification labels for the corpus.

2.8. Problems encountered

As indicated in the introduction, large databases usually pose some problems with transcription. In case of Russian, we had to face extra challenges related to lack of standardized segmental and prosodic transcriptions. As far as segmental tire is concerned, Russian SAMPA was chosen. However, transcription of Russian texts strongly depends on the phonological positions of researchers, especially as regards phonological interpretations of reduced vowels, devoicing of consonants, etc. [10]. The creation of prosodic transcription tire is difficult due to the lack of not only standardized prosodic notation, but even of a workable intonation system for Russian [11]. The prosodic analysis of the corpus is aimed at designing principles of prosodic notation and intonation system in Russian.

3. SOME PRELIMINARY RESULTS OF CORPUS STATISTICS

3.1. The duration of dialogues

The analysis of dialogue durations shows that the average duration is 8.2 minutes. Maximum duration is 22 minutes, and minimum duration is 3 minutes. The duration is longer in cases when an error in performing the task was noticed and corrected by participants. Since each subject participated in four recordings (twice as the Guide and twice as the Traveler) s/he developed game skills, which caused duration reduction of subsequent dialogues up to 30%.

3.2. The speech acts

A preliminary analysis of the corpus allowed the authors to determine the following speech acts, which are further used for the discoursal structure labeling:

1) Pre-and post-task acts:

Introduction (the Guide and the Traveler introduce each other and the task, define roles)

Conclusion (the Guide or the Traveler state the completion of the task)

2) Task acts:

Instruction (the Guide tells the Traveler to undertake some action, like finding an object on the map, drawing a line, etc.).

Explanation (the Guide states information about the location or characteristics of the landmarks)

Clarification (repeated explanation with further details)

Follow-up check (the Guide asks a question to elicit a confirmation that the Traveler follows him/her)

Align (the Guide checks whether the Traveler is ready to pass over to the next stage of the task)

Info-quest (y/n, wh, or, tag) (the Traveler requests some specific information in the form of a yes/no, special, alternative or disjunctive question)

Confirmatory question (the Traveler asks a question to confirm that what he does/thinks is right)

Echo-question (a repeated question when the partner's words are not heard/understood)

Acknowledgement (a brief comment to acknowledge correct understanding/move)

Reply (information given in response to a question)

3) Extra-task acts

Deviation (one of the partners starts talking about something unrelated to the map task topic)

Re-setting (one of the partners brings the conversation back to the map task).

4. THE APPLICATIONS OF MAP TASK CORPUS OF SPOKEN RUSSIAN

4.1. Current linguistic applications

The materials of Map task corpus are employed for constructing the intonation system of spoken Russian and testing the prosodic notation system for Russian. The materials are also used for investigating syntax-prosody interactions: the expression of sentence type and pragmatic type of utterances by syntactic, lexical and prosodic means.

4.2. Current language teaching applications

Extracts from the corpus are currently used for teaching Russian as the foreign language to Japanese university students, e.g., for teaching assimilation and sentence type intonation.

4.3. Other possible applications

The corpus can serve a variety of purposes including interdisciplinary studies of spoken language discourse.
In particular, the corpus can be used for psychological and psycholinguistic studies of diagrammatic reasoning, (collaborative) problem and error-solving strategies, taking/transferring initiative, encouraging, etc.
The corpus can be also employed for the studies of emotion expression in natural speech, as well as for a variety of speech processing and recognition tasks.

5. CONCLUSION

Map task corpus of spoken Russian is designed as a multi-purpose database serving the interests and needs of linguists, applied linguists, engineers and teachers of Russian as a foreign language. In this way it encourages interdisciplinary interactions focused on conversational speech. The corpus can be used for a wide range of experimental studies of the sound system of conversation Russian, including assimilation, reduction, elision, stress, accents, and intonation. It also provides materials for investigating all levels of conversational Russian language, such as morphophonological phenomena, lexis, grammar, syntax, discourse strategies.

REFERENCES

[1] Anderson, A. H., Bader, M., Gurman Bard, E., Boyle, E., Doherty, G., Garrod, S, Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S. 1991. The HCRC Map task corpus. *Language and Speech*, 34 (4), 351-366.

[2] Bard, E.G, Sotillo, C., Anderson, A.H., Taylor, M.M. 1995.The DCIEM Map Task corpus: Spontaneous dialogue under sleep deprivation and drug treatment. In: Isabel Trancoso and Roger Moore (Eds), *Proceedings of the ESCA-NATO Tutorial and Workshop on Speech under Stress*, 25-28, Lisbon, September.

[3] Barras, C., Geoffrois, E., Wu, Z. & Liberman, M. 2000. Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication*, 33 (1-2), and http://www.ldc.upenn. edu/mirror/Transcriber/

[4] Burger, S., MacLaren, V., Yu H. 2002. The ISL Meeting Corpus: The impact of meeting type on speech style. *ICSLP 2002*, 301-304.

[5] Caseiro, D., Silva, F.M., Trancoso, I. 2002. Automatic alignment of map task dialogues using WFSTs. In: *Pronunciation Modeling and Lexicon Adaptation for Spoken Language (PMLA)*. ISCA Tutorial and research Workshop of ICSLP 2002 (Interspeech 2002). CD-rom.

[6] Demuynck, K., Laureys, T. & Gillis, S. 2002. Automatic generation of phonetic transcriptions for large speech corpora. *ICSLP 2002*, 333- 336.

[7] Goubanova, O. V. 2002. Forms of introduction in map task dialogues: case of L2 Russian speakers. *ICSLP 2002*, 845-848.

[8] Keating, P. A. 1998. Word-level phonetic variation in large speech corpora. *ZAS Papers in Linguistics*, 11, 35-50.

[9] Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A, Den, Y. 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language and Speech*, 41 (3-4), 295-321.

[10] Makarova, V. 2002. Prosodic feature perception by human subjects. *9th Australian International Conference on Speech Science and Technology*. Melbourne, Dec 2-5, 2002. CD-rom.

[11] Makarova, V. 1998. Phonetic studies in Russia. *Goken-Forum, Journal of the Institute of Foreign Languages*, Waseda University, 9, 197-211.