

# Speech Data Corpus for Verbal Intelligence Estimation

Kseniya Zablotskaya<sup>1</sup>, Steffen Walter<sup>2</sup>, Wolfgang Minker<sup>3</sup>

<sup>1,3</sup> Institute of Information Technology, University of Ulm, Germany

<sup>2</sup> Department of Medical Psychology, University of Ulm, Germany

{kseniya.zablotskaya, steffen.walter, wolfgang.minker}@uni-ulm.de

## Abstract

The goal of our research is the development of algorithms for automatic estimation of a person's verbal intelligence based on the analysis of transcribed spoken utterances. In this paper we present the corpus of German native speakers' monologues and dialogues about the same topics collected at the University of Ulm, Germany. The monologues were descriptions of two short films; the dialogues were discussions about problems of German education. The data corpus contains the verbal intelligence quotients of each speaker, which were measured with the Hamburg Wechsler Intelligence Test for Adults. In this paper we describe our corpus, why we decided to create it, and how it was collected. We also describe some approaches which can be applied to the transcribed spoken utterances for extraction of different features which could have a correlation with a person's verbal intelligence. The data corpus consists of 71 monologues and 30 dialogues (about 10 hours of audio data).

## 1. Introduction

In this paper we present the corpus for verbal intelligence estimation. The purpose of our research is the development of algorithms for automatic estimation of a person's verbal intelligence - in other words the level of verbal cognitive processes - based on the analysis of transcribed spoken utterances. There are many psychological researches which show that different words of everyday speech may reflect psychological states of a speaker. According to these approaches the analysis of words can tell us about the sex, age and social class of a speaker. It also potentially can tell us if a speaker is telling the truth, if he has dominance in a conversation, if he has depression and much more (Tausczik and Pennebaker, in press).

We try to understand how people with different verbal intelligence interact with each other, which words and sentence structure they use, and what the differences between their languages are, when they talk about the same topic.

In our opinion, this research is necessary to develop an exact companion system technology in human-machine interaction. We are aware of the critical ethical issue of this technology, but only when a companion system recognizes the verbal ability of a user, it can precisely adapt to the user specific strategies. Future dialogue systems may use information about the estimated cognitive processes of a user and change the level of the interaction, make the interaction for a non-experienced user easier and clearer, may help the user solve different problems more effectively.

Language is the most common way for people to express their thoughts, emotions and feelings. Spoken utterances of a particular person reflect his educational and cultural background, psychology, life-experience, reasoning, his level of intelligence, etc. When we hear people speech, we can approximately estimate levels of intelligence using our intuition and perception.

In our research we try to understand how someone's spontaneous vocabulary reflects his verbal intelligence. Applying different psychological, syntactical, semantic, statistical, and lexical approaches to transcribed speech, it is possible to find and extract language use characteristics related

to verbal intelligence. As a result we are going to create a system which would be able to estimate the verbal intelligence levels of speakers. For our research we need a large amount of German utterance transcripts representing an extended and uninterrupted speech by a single person in a real-life situation (monologues).

But if we want to compare spoken utterances of people with different cognitive processes, these people have to discuss the same theme; the spoken utterances have to be about the same topic. All people have different interests: one can be an expert in technology, the other loves art, the third one is fond of cooking. If these persons discuss their favourite topics, they may use very specific vocabulary compared with other people who don't take an interest in those areas. Such topics cannot be compared because the results will not be objective. That's why we need to compare speakers' utterances about the same topic from our every-day life. So we therefore decided to record the speech of different people when they talk about the same short film and describe it with their own words.

Another purpose of our research is the analysis of dialogues. We are going to create some models for automatic estimation of verbal intelligence of dialogue participants. This task is more difficult because the speaker's level of the dialogue depends on the cognitive processes of his dialogue partner. For example, when an adult with the certain level of verbal intelligence talks to his child, his words and sentence structures are easier than those he uses when he talks to his friends or his boss. For this task we need conversations or discussions between two people about the same topic. But we can get interesting discussions only if the dialogue partners know a lot about this topic and have their own opinions. We then decided to choose the school system and education in Germany as a topic for our dialogues.

To know the levels of speakers' verbal intelligence we used the verbal part of the Hamburg Wechsler Intelligence Test for Adults (Wechsler, 1982). Using this intelligence test, the verbal intelligence quotient for each speaker can be calculated and can be used for further investigations.

In this paper we present the corpus of transcribed mono-

logues and dialogues of German native speakers and their levels of verbal intelligence. The corpus was collected at the University of Ulm, Germany.

## 2. Method

### 2.1. Films

At first we decided to collect some monologues from different speakers and to determine if it was worth continuing this research. The first 14 candidates were students and Ph.D. students from the University of Ulm with high levels of verbal intelligence. They were shown two short films from TV-program "Galileo" (Pro7, 2008). After the first film had been shown, the participants were asked to imagine that they met their good friends and wanted to tell them about one interesting film. They were asked to describe the story to their friends. Then they were shown the second film and their speech was also recorded.

The first film was about the craziest hotels in the world: a "capsule hotel" in Tokyo, where each room consists of a horizontal plastic box about 6 feet long, 2 feet wide and 2 feet high, complete with television and radio; a temporary ice hotel made up entirely of snow and sculpted blocks of ice; a Berlin theme hotel, where 40 themed rooms have a couple of real standouts, like the Flying Bed and Grandma's; different tree-houses, which are perched 8 to 10 meters above the ground, accommodate four to six people, and can be rented for the night; the Hotel Everland - the first mobile hotel and a contemporary artwork, which was installed on the roof top of the "Palais de Tokyo" in Paris in 2007.

The second film was about an experiment on how long people could stay awake. Two men and one woman were asked to stay in the same house and to fight against sleep. When they were in a bathroom, they had to sing a song or to whistle. The participants also had to take different tests to control their concentration, memory, attention, condition and a general well-being. As a result the woman won. She could be without sleep for 58 hours. At the end of the film it was told that sleep was very necessary and experiments with animals showed that being without sleep can be dangerous to your life. When we asked our participants to tell us about these films, we didn't want to get official or scientific speech. We needed normal speech, which these persons used every day when they talked to their family, children and friends, and when they talked about general things. It wasn't important how many details of these films they could remember. The main idea was to give them the same topic to talk about.

Then participants were asked to take an intelligence test, upon conditions that their names would be anonymous. Intelligence has been studied from many perspectives throughout history. The most widely accepted theory of intelligence is based on an intelligence quotient (IQ). In our research we used the Hamburg Wechsler Intelligence Test for Adults (HAWIE) (Wechsler, 1982). It's the German version of the American original. Its scale is based on a projection of the subject's measured rank on the Gaussian bell curve with a center value (average IQ) of 100 and a standard deviation of 15. The test is organized for adults ranging in age from 16 to 74 years and consists of 6 verbal

and 5 performance tests. Education, experience and life-style also contribute to scoring better on this test. For the research we used only the verbal section:

- Information. With this sub-test the general knowledge is measured; 25 questions come from a particular culture. For example, "What is the capital of Russia?"
- Comprehension. This sub-test measures social awareness and common-sense. It focuses on the social sense and the conception of cultural values. For example, "What would you do if you lost your way in a forest?"
- Digit Span. The auditory short memory, concentration and attention are measured with this sub-test. A participant is asked to repeat strings of digits forward and then backward.
- Arithmetic. Arithmetic problems are offered in a story-telling way to identify mental alertness. It focuses upon attention and concentration while manipulating mental mathematical problems. For example, "Seven envelopes cost twenty five cents. How many envelopes can you buy if you have one dollar?"
- Similarities in Dissimilar Objects. A test taker is asked to find abstract similarities among different objects, for example among "a dog" and "a lion". With this test, abstract reasoning and power of conceptualization are measured.
- Vocabulary. A participant is asked to explain the meaning of different words, for example "to crawl" or "a needle". The sub-test measures the comprehension of meanings and relations between the expressive words. For example, "What does the word zebra mean?"

After 14 candidates had been tested, our corpus contained 28 audio-files, for a total count of 10395 words.

When we asked the participants whether it was difficult for them to speak into a microphone, they said that it was a little bit unusual. Only 4 of 14 candidates said that the first film was easier to describe. The other 10 participants used restrained language talking about the hotels, but said very informative stories related to the second film. As they explained, it was difficult to remember the order of the hotels and the hotels themselves. But they could describe the second film without certain details and numbers. Then we decided to show other candidates only the second film and to use these monologues for the analysis. We want to create algorithms and models which will automatically estimate the person's verbal intelligent regardless of the theme speakers are talking about. The monologues about the first film will be used to check our algorithms and models. Examples of the monologues are shown in Figure 1 and Figure 2<sup>1</sup>.

For finding other candidates, our experiment was advertised in the newspaper. In all the experiments, we obtained permission from candidates to use the data for investigation

---

<sup>1</sup>As the conversation language is German, the example was directly translated into English.

"Well, the film was about hotels, mad hotels, not usual hotels. The first one was in Tokyo, it was a hotel where you could see only small boxes for sleeping, small beds, and there were no rooms there. It had communal showers and so on. There was also an ice hotel in the Alps. Igloos were built in the same way as Eskimos did. People slept in sleeping bags on coats. There was a hotel in Berlin, where the rooms were unusually furnished. For example, a mine tunnel, a prison cell, a castle. Exactly as everybody likes. There was a hotel which had only one room, quite small and mobile. It stands at the moment in Paris on a museum. It could stand in other places, where you would like. And there was another hotel. Yes, tree-house hotel, which stayed somewhere in a forest. Five tree-houses were built, put very high, and set in the forest. It was for people who would like to spend the nights in the nature."

Figure 1: Excerpt from one of the recorded monologues.

Hi. Yesterday I watched a film. It was Galileo. And it was about a project, about sleep, how long people could stay awake. When they were not concentrated, they had to participate in different experiments. For example, to park a car. And it was really interesting. There were three candidates and they had to perform these tasks. When they were overtired, they were getting cold. But it was 23 degrees in the room. It was interesting to know that some experiments were conducted with mice. The mice were not allowed to sleep. They died after two weeks because they were too overtired. They also said that if you are awake for so many hours, you will feel as you have drunk a glass of wine. So, it is also dangerous.

Figure 2: Excerpt from one of the recorded monologues.

and to publish results. For their participation they received 10 Euro. This time participants were also asked to make a 10-minute dialogue with a dialogue partner. The topic of the dialogue was about the education and the school system in Germany. The participants had to express their opinions, to determine advantages and disadvantages of the school system, to talk about teachers, lectures, marks and etc. For our experiment the participants could come alone or with a dialogue partner. If they intended to come alone, we asked another participant or some of our colleagues to join them to discuss this problem.

If the candidates hadn't met each other before and had difficulties in making a dialogue, they were asked to dispute and to prove the certain position about the school system. For example, they were asked to imagine that they had different points of view about German education. The first participant was asked to prove that the school system in Germany is very good, that the children get a very good education and it is no use making changes to it. The second participant was asked to describe bad features of German education, make different examples and to offer some innovations. Sometimes it was easier for the participants to dispute, because they could analyze the position of the dialogue partner and to react in some way. But sometimes it was more difficult, because the participants couldn't find

(for example) good features of the education if their private opinion was different. An example of the dialogues is shown in Figure 3.

P1/ I think that teachers work very hard. They have their lessons, but they have to prepare something for them. And after the lessons they have to check something. It takes much time. I think, they are paid for these hours.

P2/ Hmm.

P1/ So, the children have to go home and to learn their lessons with their parents.

P2/ Yes, it is very often.

P1/ But, it doesn't work!

P2/ Yes, it is not possible.

P1/ Because their parents are at work!

P2/ Yes.

P1/ Because they have to earn a living, their children need money.

P2/ Yes, the parents are overbusy.

P1/ Do you have children?

P2/ Yes, I have a son. When he went to school I saw that he wasn't overbusy, he didn't have much homework.

P1/ They have to do much homework! It is better than to play computer games. And the lessons have to be more interesting.

P2/ And they have to learn for themselves.

P1/ Yes, you are right.

Figure 3: Excerpt from one of the recorded dialogues.

Sometimes when participants couldn't find new ideas and continue the dialogue, we had to say "Thank you very much. You may end your dialogue". At that moment the participants were beginning to relax and to discuss German education or something else as if they knew each other very well. This dialogue was more real, because they were not already stressed. We had to record their speech once again. As a result we decided to record the whole experiment from beginning to end, not to miss some interesting information. During collecting the corpus, some participants were asked to make a dialogue with different partners. That's why we have an opportunity to compare the levels of these dialogues and to analyze the difference between spoken features of the same speaker in different dialogues.

## 2.2. Corpus Structure

The corpus has the following structure. All monologues and dialogues of each speaker were transcribed according to the Transcription Standards by Mergenthaler. The transcripts that are made by this set of rules are very good for different scientific researches, especially in psychology. Usually spoken utterances don't comply with grammatical rules. The punctuation marks in transcripts are used to show rhythmical and syntactical speech interruptions.

- "?" is used for interrogative word intonation and for rising tone.
- "." is used to mark the completed thought and falling tone.

- “;” is used to mark a short pause in the speech, but with a continuation of the main idea.
- “;” is used to mark an interrupted thought. For instance “no no. or yes? you say; understand;”.

Non-verbal and paraverbal aspects of the speech, repeated words are also marked.

The data corpus contains result tables of the candidates (Table 1). The table contains candidate’s points for each verbal task and the verbal IQ. The verbal IQ was measured according to the special tables of the HAWIE.

Sub-tests	Points
Information	18 out of 25
Comprehension	19 out of 20
Digit Span	15 out of 17
Arithmetic	13 out of 14
Similarities in Dissimilar Objects	24 out of 24
Vocabulary	73 out of 84
Verbal IQ 122	

Table 1: A result table of a candidate.

### 2.3. Participants

Overall, 56 candidates have been tested, 71 monologues (3 hours 30 minutes) and 30 dialogues (6 hours 30 minutes) have been collected. We are going to test other candidates and to collect more monologues and dialogues. And the first 14 candidates, who described the two films, will be also asked to make dialogues.

## 3. Approaches

Many different approaches can be applied to the collected data. Informative features can be extracted by looking at word usage, abstracts, and emotion words, exploring the broader meaning of language within a phrase or sentence, conversational turn, or an entire narrative (Walter, 2008; Mergenthaler, 1996). The transcribed speech can be analyzed at different linguistic levels: morphology, lexicology, syntax, semantics, and discourse (von der Brueck et al., 2008). In trying to describe the same film, participants expressed themselves in different ways. There is also linguistic style reflected by use of function words - articles, auxiliary verbs, prepositions, and pronouns. It shows how people talk about the given topic rather than what they are talking about. The different ways participants put their words together and express themselves describing the same film represent linguistic style (Campbell and Pennebaker, 2003). We will not go into detail here about the approaches we are going to apply to this corpus, because it is the theme of separated research.

The collected corpus can be also used to measure the degree of speakers’ immersion in monologues and dialogues. The more immersed a candidate is, the better his story sounds and a better analysis can be made. We can assess how individuals are referring to each other, especially if they haven’t met before, and how it influences a conversation. Using this speech data, gender and age differences in language use can

be analyzed. We can try to automatically estimate to what degree the collected monologues correspond to the criteria of a ‘good story’ and to use for a model training only well-made texts. Contra-dialogues can be used to check linguistic criteria whether the dialogue participants are honest and agree with the point of view they try to prove. The relative use of first person singular pronouns is an indicator of the status of two people in a conversation. We can check if the participants are engaged in the conversation and if a person with a higher level of verbal intelligence has a higher status in a dialogue. The levels of agreement can be measured. We think that this corpus is also interesting for researchers who investigate person’s speech and behavior. People of different ages and educational levels discuss the same topics. They express different emotions: they are glad, confused, hesitating, they try to keep the conversation going and to prove their opinions. They try to communicate with people which they have never seen before. We haven’t finished collecting the corpus and we are going to find some more candidates and to continue these recordings.

## Acknowledgements

This work is partly supported by the DAAD (German Academic Exchange Service).

Acknowledgements Parts of the research described in this article will be supported by the Transregional Collaborative Research Centre SFB/TRR 62 “Companion- Technology for Cognitive Technical Systems” funded by the German Research Foundation (DFG).

## 4. References

- R. S. Campbell and J.W. Pennebaker. 2003. The secret life of pronouns: Flexibility in writing style and physical health. *Psychological science*, 14(1):60–65.
- E. Mergenthaler. 1996. Emotion-abstraction patterns in verbatim protocols: A new way of describing psychotherapeutic processes. *Journal of Consulting and Clinical Psychology*, 6(64).
- Pro7. 2008. <http://www.prosieben.de/tv/galileo/>.
- Y. R. Tausczik and J.W. Pennebaker. in press. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*.
- T. von der Brueck, H. Helbig, and J. Leveling. 2008. The readability checker delite. Technical Report 345-5, Fakultät fuer Mathematik und Informatik, FernUniversitaet in Hagen.
- S. Walter. 2008. *Therapeutische Gesprachsfuehrung und Emotions-Abstraktionsmuster des Patienten, Dissertation zur Erlangung des Doktorgrades der Humanbiologie der Medizinischen Fakultät der Universitaet Ulm*. Ph.D. thesis, University of Ulm.
- D. Wechsler. 1982. *Handanweisung zum Hamburg-Wechsler-Intelligenztest fuer Erwachsene (HAWIE)*. Separatdr., Bern; Stuttgart; Wien, Huber.