# Developing and evaluating an emergency scenario dialogue corpus

**Jolanta Bachan**
Institute of Linguistics, Adam Mickiewicz University
Al. Niepodległości 4, 61-874 Poznań
E-mail: jolabachan@gmail.com

**Abstract**

The present paper describes the development and evaluation of the Polish emergency dialogue corpus recorded for studying alignment phenomena in stress scenarios. The challenge is that emergency dialogues are more complex on many levels than standard information negotiation dialogues, different resources are needed for differential investigation, and resources for this kind of corpus are rare. Currently there is no comparable corpus for Polish. In the present context, alignment is meant as adaptation on the syntactic, semantic and pragmatic levels of communication between the two interlocutors, including choice of similar lexical items and speaking style. Four different dialogue scenarios were arranged and prompt speech material was created. Two maps for the map-tasks and one emergency diapix were design to prompt semi-spontaneous dialogues simulating stress and natural communicative situations. The dialogue corpus was recorded taking into account the public character of conversations in the emergency setting.
The linguistic study of alignment in this kind of dialogue made it possible to design and implement a prototype of a Polish adaptive dialogue system to support stress scenario communication (not described in this paper).

**Keywords:** dialogue corpus design, emergency dialogue, stress scenario, speech prompt material

## 1. Aim and background

The present corpus was recorded for studying alignment phenomena in stress scenarios as a basis for implementing a Polish adaptive dialogue system to support stress scenario communication. The challenge is that emergency dialogues are more complex on many levels than standard information negotiation dialogues, different resources are needed for differential investigation, and resources for this kind of corpus are rare. Currently there is no comparable corpus for Polish. This paper summarises design criteria and resource construction procedures, and summarises the evaluation of scenario variants.

In recent years new aspects of communication have been investigated which are relevant for developing natural human-computer dialogue interaction, including alignment in communication form and content between the interlocutors (Pickering & Garrod 2004) and accommodation of interlocutors to each other (Giles et al. 1992). It has been noticed that while communicating, interlocutors tend to adapt to each other's behaviour, especially speech style, vocabulary, gestures.

In the present context, alignment is meant as adaptation on the syntactic, semantic and pragmatic levels of communication between the two interlocutors, including choice of similar lexical items and speaking style. The form, content and degree of alignment depend on the formality of the communication situation and status relations between the interlocutors. An essential distinction for emergency scenarios is between public and private situations. In public situations, interlocutors generally do not know each other and the degree of alignment of their behaviours has been found to be smaller than in face-to-face conversations between close friends (Batliner et al. 2008). In fact, there may be deliberate non-alignment between call-centre operators and stressed callers, in order to calm the caller (Bachan 2011).

For the present research, two types of dialogues were recorded in laboratory conditions: a map task dialogue and a picture description dialogue, the 'diapix task' (Bradlow et al. 2007; Baker & Hazan 2009). The map and diapix tasks are sources of semi-spontaenous speech. Both dialogues are directed at crisis situations and communication in a public setting, focussing on people who do not know each other.

The analysis of the corpus was used to design and implement a prototype dialogue system which combined text input with speech output. Its core was based on two linked finite state transducers (FST): one for the dialogue manager and one for map traversal.

## 2. Corpus design

### 2.1 Task and material design criteria

The speech prompt material was intended to invoke stress. The idea was to simulate a telephone conversation which could happen in a crisis situation. Additionally, neutral prompt speech material was added to create a corpus of control dialogues.

1. *Map task*:
    1. Emergency scenario: Subject A, the 'caller', has to instruct subject B, the 'controller' to get to a place where a man with a heart attack is waiting for help. Subject A gets a map with a marked route on it which leads past different landmarks. Subject B gets a map with landmarks only. The interlocutors cannot see each other. The task is for A to describe a route to guide B, controlling an ambulance, to the emergency location. Subject A gets a description of the situation underlining the tragic situation, tending to invoke emotions such as fear or sadness. The landmarks on the two maps differ slightly, creating communication stress; on the route of the

ambulance there are obstacles such as an accident, a traffic jam, road works, school race, which A does not know about. The stress level is further boosted by giving A a 5 minute limit for the task.

2. Neutral scenario: Subject A has to guide subject B along the streets to a cinema. The maps are almost identical: the starting point is marked on the map of subject B, and the final point is marked on the map of subject A. There is no time limit for the task.

2. *Diapix task*: The task consists of the description of a picture in order to look for differences. There is no role asymmetry: both interlocutors have the same status. Their task is to describe the picture and find the differences between them. The interlocutors cannot see each other.
   1. Emergency scenario – the picture presents an accident site. The subjects get a 5-minute time limit to finish the task in order to raise the stress level.
   2. Neutral situation – the picture of a shopping area of a town.
3. *Reading*: The task is to read a text in a neutral style.

To reduce the familiarisation with the task, it was decided to carry out the emergency scenarios first, then to proceed to the neutral tasks. Reading was recorded at the end of the session.

### 2.2 Subjects

Subjects who did not know each other or were at different positions in an academic project were selected in order to ensure that the dialogue would have a public character (Batliner et al. 2008). Where possible, pairs of interlocutors with a large age gap and different academic status were chosen, which could also affect the younger people's stress level. Additionally, 3 pairs of people who know each other were recorded as a control group. For the project, 15 males and 15 females were chosen and recorded in pairs: male – male, male – female, female – female. Subjects had no preparation with instructions and scenario materials. setting, focussing on people who do not know each other.

## 3. Corpus design

### 3.1 Creation of maps

For the two map tasks, four maps were created to the specifications, using standard drawing software.

In the *emergency scenario*, the map shows a hospital and an ambulance ready to set out for the patient. On the map of subject A, the route and a few obstacles which prevent the ambulance from taking the shortest route are marked. On the map of subject B neither route nor obstacles are marked. Both maps differ slightly in positions and types of the landmarks. The landmarks are typical buildings such as a cinema, school or shop, also trees, a lake and a car park. The emergency scenario maps are shown in Figure 4.

For the *neutral scenario*, two identical maps were created, differing only in initial and termination points. For B, the initial point is marked and for A, the termination point. The first step of the conversation is then to negotiate where to start. No route is marked on the map and the route to take depends on the interlocutors. Although A is the instructor, B may also suggest their own ideas of which way to take. The maps which differ only in the initial and the termination points marked on separate maps are presented in Figure 5.

### 3.2 Creation of diapixes

For the diapix task, two sets of pictures were used. The emergency pictures are actual photos arranged for the task. The neutral pictures are adopted from the diapix task created for recording the Wildcat Corpus of Native- and Foreign-Accented English (Bradlow et al. 2007).

On the emergency scenario diapix, there is a boy injured while sledging on a hill. The pictures were taken using standard Canon digital camera, and brightness was enhanced using standard image processing tools. Both pictures differ in 10 details: 5 changed items and 5 missing items. The differences between the photos are listed in Table 1. The photos are presented in Figure 6.

The diapixes for the neutral scenario adopted from Bradlow (et al. 2007) show a shopping area a town. They were processed slightly in order to change English names for the Polish names.

| | *Version A* | *Version B* |
|---|---|---|
| *Changed Items* | boy lying nearby sledge | boy lying on sledge |
| | boy keeps legs apart | boy keeps legs together |
| | blue sledge | green sledge |
| | snow green spade | red broom |
| | rockers on fence | rucksack on fence |
| | | rockers on bike |
| *Missing Items* | glasses | no glasses |
| | no girl on sledge | girl on sledge |
| | no red car | red car |
| | no bike | bike |
| | no light | light in the window |

Table 1: Difference between diapixes from the emergency scenario.

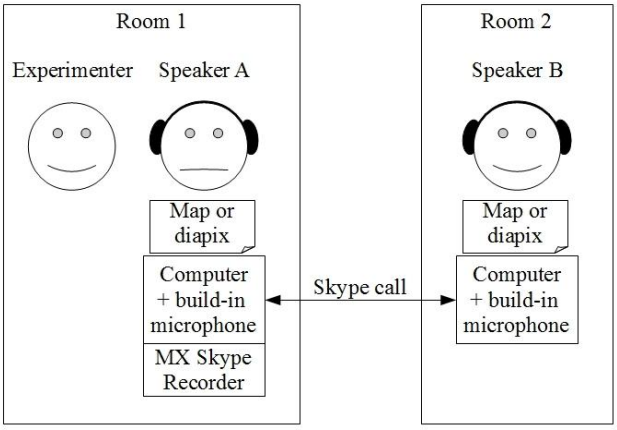


Figure 1: Recording setting of the dialogue corpus.

### 3.3 Reading task

The excerpt for the reading task was taken from the first page of "One Hundred Years of Solitude" (1967) by Gabriel García Márquez. It contains a site description and does not require any expressive poetic interpretation. The text, although has quite complex sentences, has quite simple vocabulary which should not be problematic for subjects.

### 3.4 Recording scenario

The recordings were performed in two quiet university offices. Each subjects sat alone in one room and communicated via Skype, with a laptop for Skype communication and recording. The recording setting is presented in Figure 1.

The recordings were performed by the MX Skype Recorder software on one of the laptops which allows to record unlimited time audio Skype calls on two separate channels for two speakers in the stereo WAV format. On the computer of the caller A, there was a big timer which showed the time left for the tasks. The timer was provided by the TimeLeft desktop utility distributed as a freeware and a shareware (NesterSoft Inc.).

## 4. Corpus creation

### 4.1 Corpus data

15 pairs carried out four dialogues based on the given tasks. 15 women and 15 men were arranged into 5 female-female pairs, 5 female-male pairs and 5 male-male pairs. 12 pairs were composed of people who did not know each other or were in the superior-inferior relation, and 3 control pairs were composed of friends. Into the leader's position, young students were put. Into the follower's position, people with a higher degree or in a superior position at work were put. Additionally, the reading task was performed by each of the subjects. Altogether, 60 dialogues recorded at 48kHz sampling frequency by MX Skype recorder and 30 reading tasks recorded with Praat at the 44,1kHz sampling frequency.

One recording session lasted about 30min. The corpus contains 4h 12min of speech. Detailed data of the corpus is presented in Table 2.

| | *MT: ER* | *DP: ER* | *MT: N* | *DP: N* | *Read* |
|---|---|---|---|---|---|
| *All* | 2325 | 3957 | 2353 | 5391 | 1111 |
| *Min* | 54 | 59 | 72 | 150 | 31 |
| *Max* | 325 | 310 | 251 | 568 | 49 |
| *Mean* | 155 | 264 | 157 | 359 | 37 |
| *Overall* | 15137sec = *4h 12min*<br>2726298KB = *2.6GB* | | | | |

Table 2: Data of the corpus recording (in seconds). MT - tap mask, DP - diapix, ER - emergency, N - neutral.

### 4.2 Corpus annotation

Annotation of the map data was on six tiers:

1. phones – extended Polish SAMPA phoneme set (Demenko et al. 2003); additionally, the beginnings of words and syllables were marked
2. syllables – syllables and filled-pauses
3. speech – orthographic transcription in Polish
4. English – English translation of the Polish speech tier
5. dialogue acts – Bunt's dialogue acts main categories (Bunt 2000)
6. special – on this tier speech events such as filled-pauses, confirmations or hesitations are marked.

### 4.3 General analysis of the corpus

The full corpus consists of four parts:

1. Map task: emergency – fast speech, very formal, Speaker A domination
2. Diapix: emergency – formal speech, cooperative dialogue
3. Map task: cinema – informal speech, a lot of auto-feedback coming from Speaker B, laughter, fun talk
4. Diapix: shopping area – informal speech, cooperative dialogue, diminutives

The created scenarios turned out to be very suitable settings for different kinds of dialogues. The public character of the conversations as well as the stress elicitation techniques worked as planned. In the first emergency map task dialogue, the A speakers were under stress. Their speech was fast, they were making mistakes, but aimed at finishing the task quickly. They were dominant in this task taking much of the dialogue time. On the other hand, the B speakers were calm and cooperatively non-aligned with the interlocutors. They tried to finish to the task successfully, but their speech was not affected by stress or fear. In this scenario, both speakers were very formal and used honorific forms of address.

In the diapix emergency dialogues, the A speakers were not so dominant. They started the description of their picture in order to find differences, but when their ideas finished, they let the B speakers talk. In these dialogues, the interlocutors were also very formal and their speech was affected by the stress factor.

The neutral tasks with the route description to the cinema totally changed the A speakers. Although the pairs of speakers were the same people and still they did not know each other previously, their style of speaking changed completely. First, their speech was relaxed, it was much slower and it could be noticed that they were happy with the idea of helping others with such a leisurely task as a visit to the cinema. Second, in their speech many colloquial words and phrases appeared which were not present in the previous dialogues. The speech of the A speakers was not dominant to the same extent as in the emergency scenario. Both speakers were making additional comments not connected with the task itself which could be described as 'fun talk'.

The recordings of the last scenario, the diapix of the shopping area, also differed from the other dialogues. In many cases, Speaker B was dominant, as in a normal setting where he or she would be socially superior to the student. However, these dialogues were still very cooperative, including colloquial vocabulary, diminutives and laughter.

### 4.4 Dialogue analysis on FST

One emergency map task dialogue was analysed in detail for alignment phenomena.

The emergency map in Figure 4 can be represented as a

finite state transducer (FST) where each junction corresponds to the transition node. In Figure 3 the different transitions are presented and in Figure 2 the FST is shown. However, in the emergency map not all the streets are open. Some junctions cannot be reached, because there is no way through. There is a traffic jam on the way or roadworks, and even at one place the street has been blocked because of a school race. Such junctions are not taken into account when designing the FST. Traffic on all the other streets is two-way, so turnings back are not hampered. Moving along the map, some route is followed. On a normal map, the route can be tracked thanks to the street names or the landmarks being passed on the way. In the FST, the street names and the landmarks can be replaced by Latin letters for simplification. Such an analysis of the map resulted in creation of a FST, modelling the movements along the map in order to reach the goal (see Bachan 2011 for details).

The emergency dialogue was analysed in order to find correspondence with the FST. One dialogue was divided into 29 utterance exchanges according to the specification described below. The first 2 exchanges were aimed to open the dialogue and set the topic. The last exchange aimed at closing the dialogue. These 3 utterance exchanges were not taken into account in the current analysis as they did not include instructions about how to move on the map. But it is important to note that in the second exchange when the topic was set, the speakers agreed from where to start the route, so they described the start node. The other 26 utterance exchanges were analysed and instructions were compared with the transitions of the FST. Each instruction led to moving from one node to another node of the FST, resulting in creating a route as a sequence of Latin letters. However, spoken instructions were not structured as it would be expected for the FST. Spoken instructions led to jumps over one or more nodes, neglecting the nodes which were on the way. Although unclear for the FST, the instructions were understood by the human partner. Sometimes the instructions led to jumps back in order to explain the route again. Finally, there were utterance exchanges clarifying the current position on the map.

In this dialogue, the alignment phenomenon may be observed as a smooth move along the map, i.e. the FST. As long as the alignment is not disrupted, the interlocutors exchange information – instructions and positive auto-feedback, and move from one landmark to the next landmark on the map. However, immediately when a deviation from the alignment takes place, a misalignment is reported and the interlocutors try to clarify the misunderstandings. The misalignment correction phenomenon relates to the finite state model as the backward movements in the traversal of the automaton.

The alignment on the semantic level is interpreted as obtaining the same semantic representation of the map. Because the maps which the interlocutors see differ, a misalignment must take place sooner or later. The recovery process from misalignment is very quick and intuitive and does not require many explanations from the speakers. An examples of misalignment from the semantic representation of the is (spk means speaker's noise, * is a disfluency marker):

*A: spk//it means* of course you don't you don't actually have a choice there at the roundabout yyy there are roadworks so right*

*B: aha*

*A: spk*

*B: spk//because actually at* on* on my map it says there are no roadworks.*

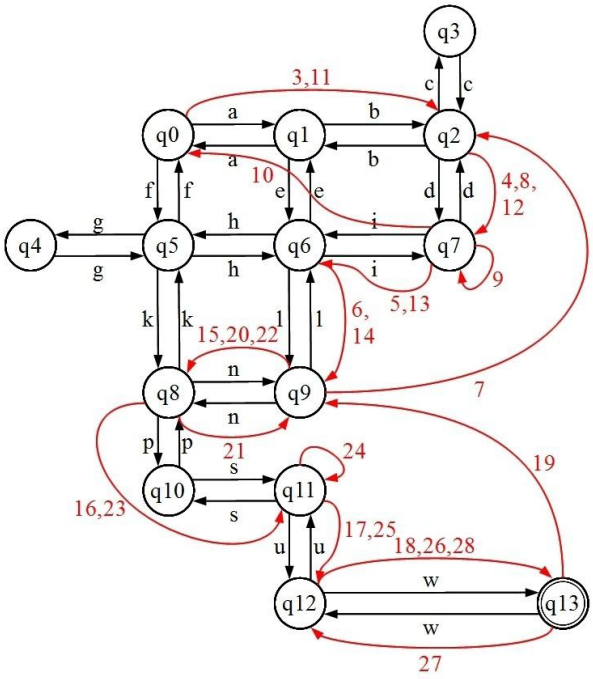


Figure 2: Map FST with utterance exchanges IDs

The dialogue chunks about transitions were transferred onto the map FST. The instructions in the utterance exchanges led to moving along the FST and these moves are visualised as curved arcs with utterance exchange IDs in Figure 2. Following the utterance exchange IDs shows how the dialogue proceeded. Analysis of the data about the transitions and utterances and dialogue act categories shows the following characteristics of the dialogue:

1. Most of the instructions *forward* expressed in one dialogue utterance exchange led to move forward from one node to adjacent node. Only the instruction leading from q0 to q2 nodes (IDs: 3,11), did not explicitly underline the transition at q1 node. Also the instruction which led to go around the roundabout did not underline the intermediate q10 node (IDs: 16, 23).
2. The turnings back (loops) lead to moving far *backward* over a few nodes (IDs: 7, 10, 19). There are turnings back to the adjacent nodes (IDs: 21, 27) but other go backward over up to 4 nodes (ID: 19).
3. Every move *forward* is repeated, therefore the are always at least 2 different IDs over each forward arc. This means that the instruction forward were repeated to make sure the mistake is avoided and the move is correct.
4. Utterance exchanges could be visualised as local loops (IDs: 9, 24) and they did not lead to any move either forward or backward.
5. Each move *backward* and each local loop was initiated by information seeking dialogue act.
6. Speaker A produced many information providing dialogue acts and directives and Speaker B role was limited to giving positive auto-feedback and confirmation dialogue acts.

7. Disagreement dialogue acts by Speaker A were preceded by information providing coming from Speaker B (IDs: 8, 20). This means that whenever Speaker B was uncertain about the route, he presented his reasoning in information providing dialogue acts and this led to disagreement and Speaker A tried to explain again the route. Disagreement (ID: 8, 20) led either to a local loop (ID: 9) or a turning back (ID: 21).

## 5. Results, analyses, software implementations

The resources include:

1. a carefully designed dialogue corpus, a rich source of alignment phenomena and recoveries from misalignment.
2. a human-computer dialogue model, implemented in a prototype dialogue system, in which the human-computer dialogue is handled by two linked finite state automata: one for the dialogue manager and one for the map traversal (Bachan 2011).

Evaluation of the results of the experiment showed clearly, in general expected but sometimes unexpected differences between behaviour under different conditions.

## 6. Acknowledgements

This work was partly funded by the research supervisor project grant No. N N104 119838. The author is currently supported by grant "Collecting and processing of the verbal information in military systems for crime and terrorism prevention and control". (OR 00017012).

## 7. References

Bachan, J. (2011). Modelling semantic alignment in emergency dialogue. In *Proceedings of 5th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*. 25-27 November 2011, Poznań, Poland, pp. 324--328.

Bachan, J. (2011). *Communicative Alignment of Synthetic Speech*. Ph.D. Thesis. Institute of Linguistics. Adam Mickiewicz University. Poznań, Poland.

Baker, R. & Hazan, V. (2009). Acoustic-phonetic characteristics of naturally-elicited clear speech in British English. (A) In: *J. Acoust. Soc. Am., 125*, pp. 2729.

Batliner, A., Steidl, S., Hacker, Ch. & Nöth, E. (2008). Private emotions versus social interaction: a data-driven approach towards analysing emotion in speech. In: *User Modelling and User-Adapted Interaction - The Journal of Personalization Research 18*, pp. 175--206.

Bradlow, A. R., Baker, R. E., Choi, A., Kim, M. and van Engen, K. J. (2007). The Wildcat Corpus of Native- and Foreign-Accented English. In: *Journal of the Acoustical Society of America, 121(5)*, Pt.2, pp. 3072.

Bunt, H. (2000). Dialogue pragmatics and context specification. In H. Bunt & W. Black, (Eds.) *Abduction, Belief and Context in Dialogue. Studies in Computational Pragmatics*. John Benjamins, Amsterdam, pp. 81--150.

Demenko, G., Wypych, M. & Baranowska, E. (2003). Implementation of Grapheme-to-Phoneme Rules and Extended SAMPA Alphabet in Polish Text-to-Speech Synthesis. In G. Demenko & M. Karpiński (Eds.) *Speech and Language Technology, Vol. 7*. Poznań: Polish Phonetic Association, pp. 79--95.

García Márquez, G. (1967). *Cien anos de soledad*. Polish: *Sto lat samotności*. Translated by Grażyna Grudzińska. Warszawskie Wydawnictwo Literackie Muza SA. Warszawa 2004.

Giles, H., Coupland, N., & Coupland, J. (1992). Accomodation theory: Communication, context and consequences. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation* (pp. 1--68). Cambridge: Cambridge University Press.

MX Skype Recorder v4.3.0 Jan 30 2010. Copyright © 2006-2010.

Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. In *Behavioral and Brain Sciences, 27*, pp. 169--225.

Skype. Copyright 2003-2011 Skype Limited.

TimeLeft, Version 3.55. Copyright © 1999-2010 by NesterSoft Inc, <www.nestersoft.com/timeleft>, <http://www.timeleft.info/>, accessed 2011-01-25.

## 8. Figures

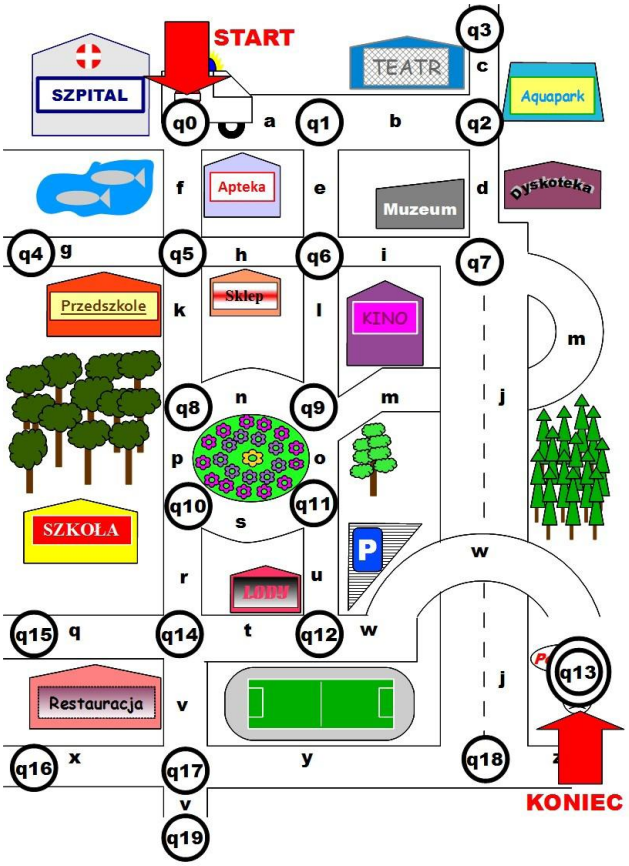


Figure 3: Map task as a basis for map traversal automaton.

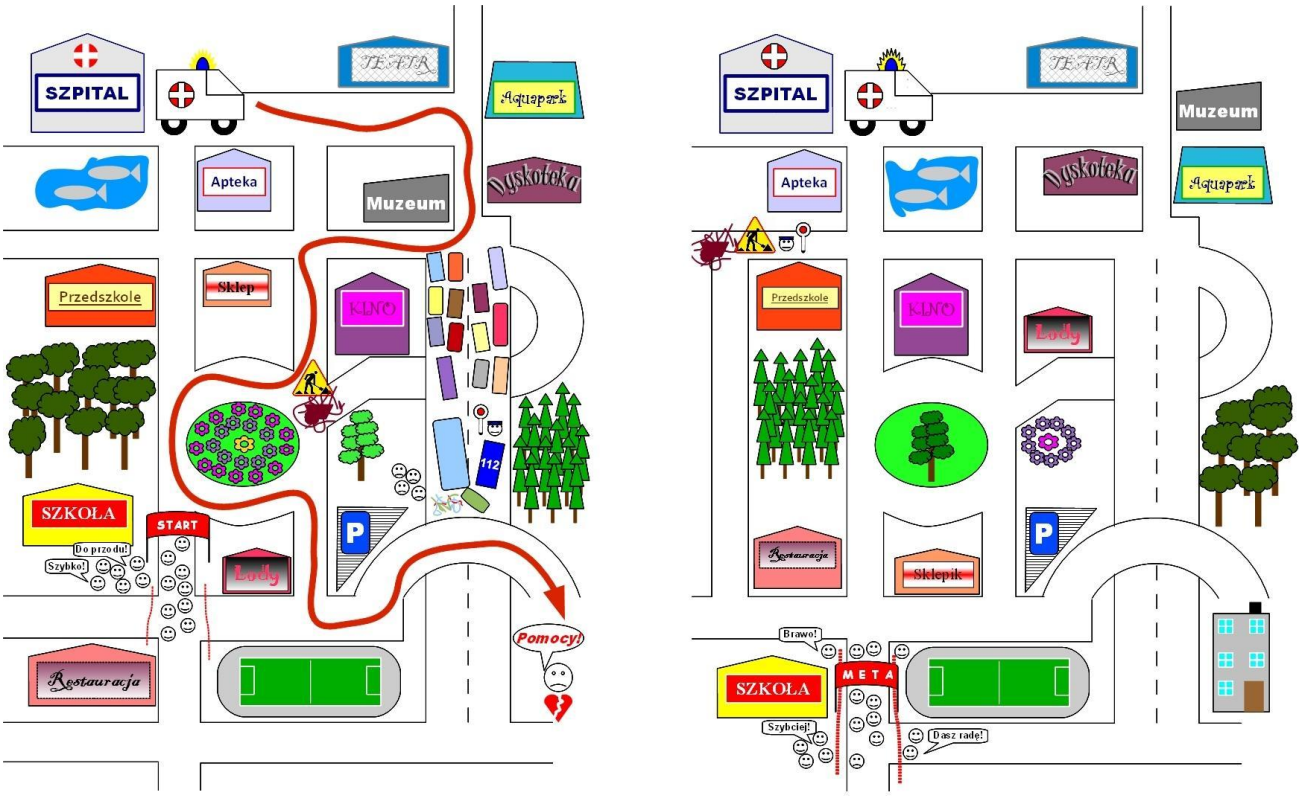


Figure 4: Emergency maps for the caller A (left) and the controller B (right).

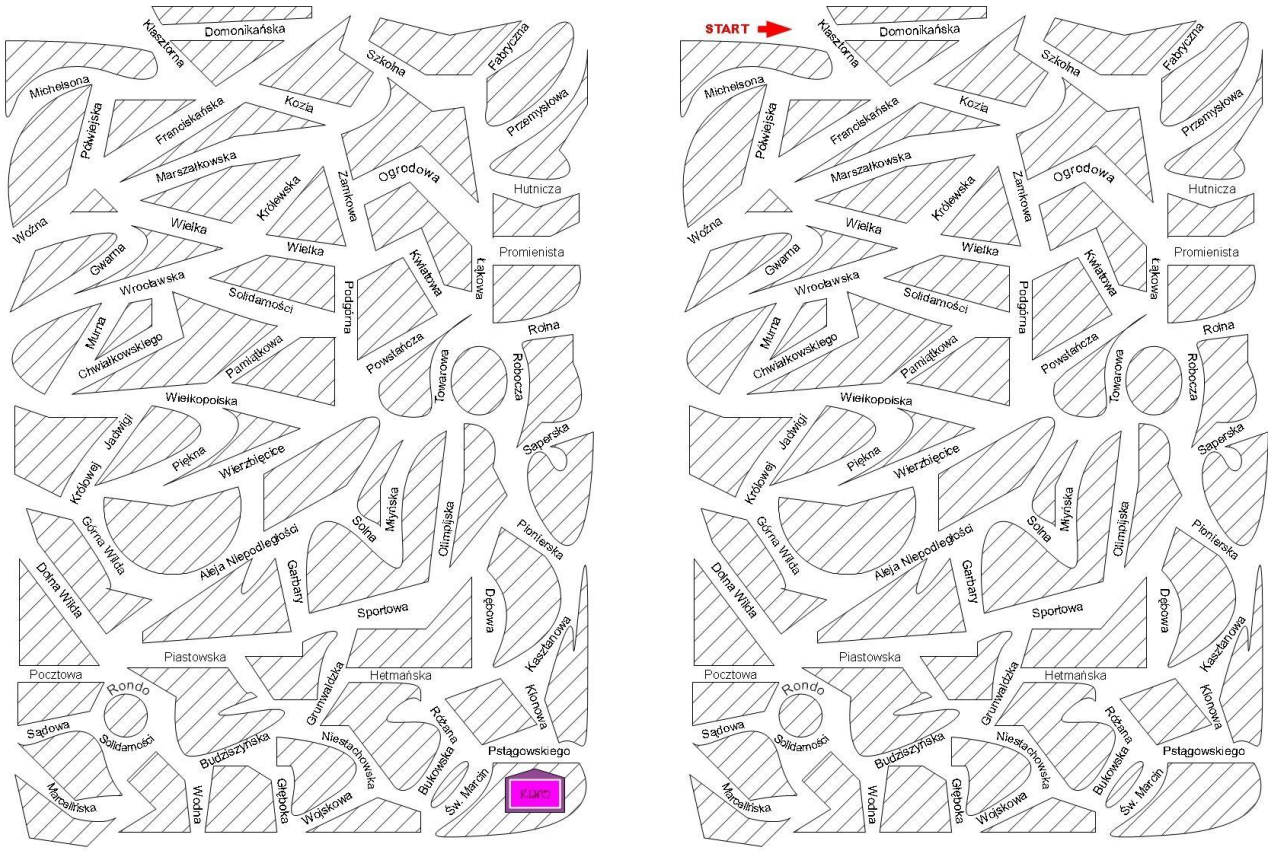


Figure 5: Neutral maps for person A (left) and person B (right).

Figure 6: Diapixes from the emergency scenario; person A (left) and person B (right).