Poznań Studies in Contemporary Linguistics 44(2), 2008, pp. 197–213

doi:10.2478/v10010-008-0010-2

POSSIBILITIES OF ANALYZING VISUAL CONDUCT WITH AN EYETRACKER DEVICE: SEARCHING FOR VISUAL DIALECTS

PÉTER BODOR[1]
ELTE University, Budapest
peter.bodor@gmail.com

ANIKÓ ILLES
ELTE University, Budapest

ABSTRACT

In this paper we will consider the possibilities of recording gazing for visual analysis. Besides describing and characterizing some general possibilities of eyetracking in visual research, we will present a design and a preliminary study which aims at analyzing an aspect of visual conduct, namely the "style" of gaze behavior in public places. Since its beginnings, sociology, just like some other social sciences, has been trying to understand the details and consequences of living in a mass society as opposed to a more traditional way of living. Specifically, on the basis of a "visual interview" procedure and with the help of an eyetracker instrument, we analyzed and compared the gaze behavior pattern of two groups – rural people and city dwellers – while in a simulated situation of walking in the street. Our preliminary results indicate that there may be differences between the gaz e behavior of little town people versus city dwellers while watching a couple of scenes recorded from a pedestrian's perspective.

KEYWORDS: visual analysis; eyetracking; visual interview; gaze behavior; viewing dialect.

0. Introduction[2]

Certain techniques which make it possible to record naturally occurring events play a crucial role in the study of human communicative behavior. For example, Sacks turned to the analysis of conversations because the tape recorder allowed the recording of talk

[1] Corresponding author.

[2] This article is an extended and upgraded version of the paper read at the 38th Poznań Linguistic Meeting, Gniezno 2007.

in interaction, and due to this, its repeated listening and sophisticated analysis (cf. Psathas 1990: 24). Similarly, when video recording became available, analysis of the establishment and organization of interaction was extended to visually available conduct (Heath 1997; Goodwin 2001). In our study, we will consider the possibilities of recording a further aspect of human conduct, namely gazing. In other words, we will ponder on the usability of eyetracker instruments – which record the timing and spacing of gaze direction – in research on interaction.

Since its beginnings, sociology, just like some other social sciences, has been trying to understand the details and consequences of living in a mass society, as opposed to some more traditional ways of living. However, relatively few studies have actually raised the issue of change in interaction conduct in this regard. In addition to describing and characterizing some general possibilities of eyetracking in research on interaction, we will present a research design and a preliminary study which aims at analyzing an aspect of visual conduct, namely the "style" of gaze behavior in public places. Specifically, on the basis of a "visual interview" procedure and with the help of an eyetracker instrument, we will analyze and compare the gaze behavior pattern of two groups of people – rural people and city dwellers – while in a situation designed to simulate walking in the street.

The ritual requirements of public behavior – in our case walking in the street – in relatively smaller communities, where some kind of premodern code of conduct dominates, include individual identification of others, such as other passers-by. At the same time, it seems plausible to us that this ritual requirement − perhaps due to ecological/ cognitive reasons − cannot be fulfilled in big towns or cities, where a more "modern" code of conduct dominates and where too many people are passing by too quickly. Our procedure intends to analyze the issue whether there are differences between the gaze behavior of small town dwellers versus city dwellers while watching a couple of scenes recorded from a pedestrian's perspective.

1. Eye tracking and visual research

Visual research is certainly a wide and rich area of study. People from such various disciplinary affiliations as art history, sociology, anthropology, linguistics and psychology all contribute to its cultivation. Furthermore, depending on its subject matter, the realm of visual research can be divided into two broad areas. On the one hand, visual research can be directed to such *man-made visual material* that people produced for whatever practical or detached reasons. These visual materials include drawings, writings, paintings, photos, diagrams and similar artefacts produced by humans, either with their bare hands, or with some instruments. Further examples for man-made visual material are shots taken by surveillance cameras, family photos, pieces of artwork, drawings of prospective patients made for the purpose of diagnosis, hand writings, traffic signs, dia-

grams, printed documents, pictograms, pictures taken by anthropologists on the field, MRI recordings, sketches of archeologists and drawings of anatomists. On the other hand, visual research can engage into analyzing any *natural visual appearances*, any visual aspect of the world, the human body and conduct, including ongoing interaction between people. Interest in natural visual appearances frequently leads to some pictorial recording of nature, the human body or its conduct respectively, thus fusing the interest in visual aspects of nature or human life with interest in visual material.[3] Nevertheless, it seems to us that it is important to differentiate between these two possibilities, at least from an analytical point of view.

In the traffic of visual material and pictures, three different, although interwoven aspects or phases can be delineated and put under scrutiny. First, visual appearances, images or films can be analyzed *in themselves*, without any specific interest in their makers or audience. Second, their *ways of producing* can be analyzed as well. Third, the beholder's *perception* and interpretation of the picture or visual appearance could be at issue. Thus, while the first type of curiosity is directed to the inherent properties of visual material, the second and third are more interested in their use.

Human vision is an intricate process, during which fast eye movements, so called *saccades* (they last from about 20 to 100 milliseconds, angular speed up to 1000 degrees per second, which is one of the fastest movements the human body is capable of performing), and *fixations* (on average, some 300 ms) alternate (Secular and Blake 1994; Bridgeman et al. 1994). Under normal circumstances, the details of this fast process are hardly visible to an observer. Practically, there is no vision during saccades, "ordinary objects are not seen during the movements" (Woodworth and Schlossberg 1963: 502). During fixations, *peripheral vision* provides general information on the visual scene, but fine details of the visual environment are perceived only with *foveal vision*, which has a relatively small area. In order to get a clear high-acuity view, the person has to turn his/her eyes so that the image of the interested area falls on the fovea.

Remote eyetrackers usually calculate the line of regard from the beams reflecting back from the cornea, and in this way the place where the person's eyes are directed to can be identified in the original picture or scene (for an overview of the recent technological developments, see Böhme et al. 2006). Eyetrackers provide a detailed *representation on visual conduct*; therefore open it up to analysis. In general terms, an eyetracker apparatus records the spatial and temporal pattern of gazing as well as the visible scene and coordinates them. It makes it possible to analyze what regions of the visual environment – in a picture or in a scene – a person's gaze is directed to and what pattern it follows in time. An eyetracker, therefore, can be used with artificial visual ma-

[3] However, this is not a logical necessity. Interposing the two interests is rather a contingent fact. In principle, visually accessible information, for example on humans, could be analyzed without any recordings of them. It seems to us that recording is rather motivated by practical reasons, especially because of the rather limited time span of our visual short term memory. Indeed, eidetic people would not need recordings for certain analysis of pictures. Some kind of recordings seem to be unavoidable on a further level of understanding, however, when the intersubjective validation is at issue.

terial and "natural" visual objects alike. The data it provides stem from the viewers' actual perceptual activity, so it is eminently suitable for the study of beholders such as readers (Rayner 1998), picture viewers (Unema et al. 2005) or film viewers (Velichkovsky et al. 2002). At the same time, since creating a visual material involves watching on the creators' side as well, their visual activity can also be detected by an eyetracker (Miall and Tchalenko 2001). It can be applied to describing both the creators' and the spectators's way of viewing any visual object. Summing up, eyetrackers have enabled us to perceive what otherwise remains imperceptible. To some degree, they could be compared to a magnifier, since they make visible what would otherwise escape the analytical eyes.

In our view, a twofold division can be drawn within the field of eye movement studies in psychology, although the two sets complement each other. The first set of studies are primarily interested in describing the characteristics and *properties of eye movements*, including the saccade-fixation patterns; they seek to explain them on a neuropsychological level and operate with psychological models (e.g. Bridgeman et al. 1994; Unema et al. 2005). A second set of studies are less interested in the fine characterization and explanation of eye movements, and particularly gazing as such, but rather consider them as behavioral measures which can help us to uncover the logic of various relevant aspects of human conduct – from language processing (Tanenhaus et al. 2000) to being a smoker (Mogg et al. 2003). Although an eyetracker system provides tremendous data which are tempting to calculate with, it is our position that this second type of research on eye movement should consider eyetracker primarily as analogous to a magnifier. Under this second stance an eyetracker instrument is a device that offers an access to certain visually guided *properties of human conduct*.

In sum, eyetracking instruments could be useful in visual analysis when the creation of and/or the use of visual material is at issue. Furthermore, it can be useful when the dynamic processes of perception of visual appearances are to be analyzed. Whenever a person actively uses her/his visual capabilities during an actual or a simulated performance, an eyetracker as an observational device, as a kind of magnifier, could offer invaluable information. In this way, at least in principle, all properties of conduct which involve active watching, including visual aspects of the organization of interaction, can be studied by eyetrackers.

2. A visual interview study: Design and initial results

2.1. A personal experience and a hypothesis on gazing styles in public places

Basic *motivating observation* or experience for the present investigation goes back to 1981. One of the authors of the present study had moved from his original residence, a little town, to a big city in order to pursue his studies. Soon after, he noticed that walk-

ing in the street, traveling by underground, particularly the use of escalators, was an utterly tiring endeavor for him, sometimes even causing bodily symptoms culminating in nausea. It took him weeks, perhaps some months, before traveling in the city ceased to be an exhausting task.

What could have been the reason for this strange experience? One can imagine that the person was trying to follow a prescribed *norm of watching* in public places which he had brought with himself from his small hometown. This norm prescribes that one must try to identify individually all passers-by and greet all acquaintances. However, this type of gazing behavior can not be maintained for a longer time if the number of oncoming people is too high. It is just too exhausting. If too many people are coming from the opposite direction we are not able to identify them individually any more, and we cannot decide whether they are our acquaintances or not. In spite of all this, people who have just moved to a big city not long ago keep on adhering to their small-town norms of watching for a while, before giving them up gradually.

On the basis of the personal experience detailed above, we formulated a hypothesis on a particular type of interactionally relevant visual conduct, namely on the varieties of gazing in public places or *visual dialects*. County people who move to the city do engage in their usual gazing conduct, and they keep actively searching for possible acquaintances. The gazing style which is characteristic in their home environment is applied under the new circumstances and persists just like a habit. Country-style gazing may be different from the city dwellers' gazing style: city people are less predisposed to search for possible acquaintances in overcrowded scenes than people living in the country.

In more technical terms, our hypothesis pertains to the organization of unfocused interaction in public places (Goffman 1963). A common-sense feature of communicative behavior in the Hungarian society, just like perhaps in all other societies, is that people who know each other, and/or want to interact with one another, greet each other if they meet in a public place. Greeting is considered as the initial move in engaging in a jointly achieved performance, in an encounter. Furthermore, this feature of the human transactions is characteristic not only of the beginning of an encounter in its strict sense; greeting is also a part of the *ritual requirements* of public behavior (Goffman 1982) if two or more people who know each other meet accidentally in a public place. On these occasions, people are not able, or might not wish, to engage in any kind of immediate encounter; rather, they just reassure each other that they have noticed each other, and reciprocally maintain their status as acquaintances. Greetings may be realized by hand, face and verbal tokens, and are considered functionally as phatic communication (Riley 2007). However, greetings presuppose recognition of the other's identity. Greetings are not aroused from vacuum. Aristotle could be right with regard to the nature of people space: *horror vacui*.[4] Greetings follow recognition. Sequentially, greetings are public acknowledgements of recognition. They ensue from recognition, and evoke an immediate

[4] The concept of "people space" was introduced by Harré and Gillett (1994: 29) as a constituent of discursive ontology.

or anticipate a postponed encounter, a focused interaction. In the case of acquaintances, this recognition is individuation. Recognition not only gives us opportunity and even command to greet the other, but at the same time, it prescribes which type of greeting is to be issued. The most usual way of recognition in a public place is based on visual perception, possibly foveal vision. If the first unit or move in a developing encounter is *greeting*, then the *individual recognition* of the other person can be numbered as zero unit. However, there is some activity before recognition as well. It is *active visual scanning* of the passers-by. Let us refer to this as the minus-one unit of an encounter. Our hypothesis on different gazing styles relates to this minus-one phase of an encounter.

We anticipate that, with the help of an eyetracker, data obtained during the visual interview method detailed below can serve as evidence for different levels of active searching. A more active visual searching directed to the identification of others would indicate a particular "premodern" gazing style. In other words, if our hypothesis is true, more frequent looking at other people's faces, and perhaps more time spent on watching them, would be characteristic of people who reside in little towns or in the country, as opposed to city dwellers.

2.2. Participants, scenes to watch and the visual interview procedure

In order to investigate the possible differences between styles of watching or dialects of viewing a scene full of people, we selected two groups of people according to their present residence. Subjects living in relatively little towns (less than 60,000 inhabitants) or villages (less than 6,000 inhabitants) were recruited for the first group, while subjects residing in big cities such as Budapest or Vienna were put in the second group. Altogether, we conducted 12 visual interviews, 6 in each group. Unfortunately, due to technical reasons, we had to exclude one person from each group. In this way, the following description of data and its analysis is based on data we obtained from 10 persons.

Although we recruited the subjects on the basis of their present residence, following the interview procedure a more detailed questionnaire was administered to all participants. The questionnaire was administered *post hoc*, since we did not intend to orientate our participants in advance, before the procedure. The questionnaire was aimed at gathering information on previous experience in living in and/or visiting big cities. In this way, we wanted to identify those country dwellers that might have had more experience in commuting to big cities versus persons who have had much less experience in this regard. Therefore, the questionnaire was especially relevant for people in the first group, who do not live in big cities, since their relevant experience (such as frequent visits to cities like Budapest, or having lived in a city before) might have an impact on their putative "country-style" gazing at crowded scenes. Note that some of the people in the country group used to attend school in Budapest, and some of them visit Budapest on a monthly basis. In this way, to a certain degree, they might get accustomed to the city-type public behavior. The relevant data on our subjects can be found in the Appendix.

The films presented to the subjects were selected from a dozen of short shots taken in the Budapest underground, particularly on escalators, while the camera was directed towards oncoming people. We used a commercial digital camera. Actually, two scenes in which the camera was moving downstairs were selected and edited into one film in such a way that a 3 second black picture with metro "noise" introduced both scenes. The film lasted 2 minutes and 32 seconds.

Subjects were told in advance that they were supposed to watch two short films while their eye movements were recorded. For the purpose of the research, they were told that we were interested in the way they watch films. The actual *visual interviews* proceeded as follows. Subjects in both groups were seated in front of a 17 inch computer display; their eyes were approximately 60 centimetres from the monitor.

In the first phase or condition, subjects were given the following instruction: "Imagine that you are traveling in the Budapest underground with a local friend. Watch the following scene in such a way!"[5] Then a calibration procedure started, which was followed by the film. While the person was watching the film, the eyetracker instrument recorded his/her eye movement.

The next phase of the experiment, i.e. the second condition, was almost identical with the first one; however, this time, the instruction was slightly different: "Imagine that you are traveling in the Budapest underground with a local friend. At the same time, try to find out whether there is any of your acquaintances on the film. Watch the following scene in such a way!" With the first instruction, we intended to capture watching habits as they might work during traveling in the metro, moving in a crowded scene. This condition will be referred to as the *"mere traveling" condition*. Nevertheless, it is highly plausible to suppose that the mental set or attitude of the person, in other words, the purpose of moving around, may influence the pattern of watching. Thus, the main reason for the second instruction was to show that watching style can be modified depending on the subject's purpose. Furthermore, watching the same films for a second time but with a slightly modified instruction made it possible to compare if changes in the watching pattern could be attributed to the different instructions. This second condition is termed as "acquaintance or friend searching", or in short the *"friend searching" condition*.

2.3. Analyzing gaze patterns and its results

The Eyegaze System we used provides data on the location of the gaze on the display (in pixel coordinates) at a 120Hz timing. The instrument we used to register eye movements is called the Eyegaze Analysis System and was developed by L. C. Technologies. Consecutive samples on direction of gaze were taken as fixation if the difference was

[5] We included the friend in the imaginary traveling in order to reduce uncertainty and searching for sign posts – a possible tendency in people traveling alone in a strange city.

less than 20 pixels. For handling the resulting data, the NYAN software (developed by The Interactive Minds, Dresden) was relied on. An output of the system is a matrix containing the sequence of timing and location of each fixation.

For a further analysis of the raw data provided by the eyetracker instrument, we developed the following procedure. NYAN also provides a possibility to replay the original film overlaid by a circle marking the location of fixations of the viewer. By replaying the films in this way, we coded the fixations. Two coding categories were devised: *face regions* versus *non-face regions*. Figure 1 and Figure 2 present examples of the coded material.

Figure 1. A Fixation on Face, 517 ms: Person no. 9, frame no. 111–116.

In this way, we complemented the time and space parameters of gazing with information on its target: face or not-face. Furthermore, the process of coding required an estimation of the size of foveal vision, where the vision is most acute. Since foveal vision is approximately 2 degrees, and the distance of display was about 60 cm, the resulting diameter of the relevant circle on the display was 2.09 cm.

In order to describe the relevant gaze pattern of the persons who attended our visual interview, we selected a 10-second sample of data from each subject's recording. In the present paper, gaze behavior from the 5th to the 15th second of the presented film will be taken into consideration for further analysis. In other words, we analyzed the number and the time of fixations in this time window. In practical terms, the analysis consisted of coding the registered fixations according to their landing positions.

Figure 2. A Fixation on Non-Face, 333ms: Person no. 9, frame no. 120–126.

Note that by operationalising our interest in public gazing into the just-described procedure of coding, we entertain the supposition indicated above that individual identification, as opposed to mere categorical identification, one mostly relies on perception of faces, and this is produced by foveal vision. According to our interview variations – watching the overcrowded scene as if merely traveling in the underground versus trying to find some acquaintances among the passers-by – we coded 2 samples for each person. Altogether, we coded 200 seconds of gaze behavior.[6] The resulting data for both interview conditions includes:

(a) the number of fixations on faces versus other areas of the scene;
(b) length of each particular fixation on faces versus non-faces from which dwell time on faces *versus* other areas was calculated.

Before presenting our results, we have to emphasize again that at the moment our analysis is based on 10 persons' recordings, and 10 seconds of watching was coded for each person per condition. For this reason, we present only descriptive statistics.

First, we will consider the pattern of gazing according to interview conditions. Our data show that both groups of people followed the instruction, and as expected, their

[6] Obviously, the hand-coding procedure we applied is rather time consuming. Further research could certainly utilize a special software, which is under construction.

gazing on faces was increased in the "friend searching" condition, as opposed to the "mere traveling" condition. The change in gaze pattern is given in proportion of overall number of fixations in Figure 3.

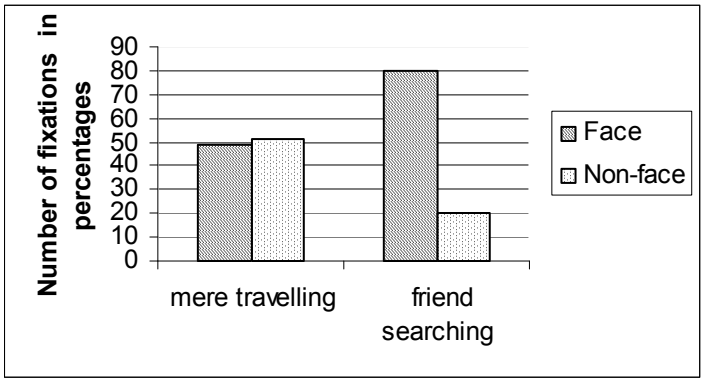

Figure 3. Number of fixations on faces and non-faces at the "mere traveling" *versus* "friend searching" condition for all participants in percentages.

As Figure 3 shows, in the "mere traveling" condition participants watched oncoming people's face roughly for as much time (49 percent) as they watched other aspects of the scene (51 percent). The situation changed dramatically when they were asked to decide whether some acquaintance or friend could be seen in the film (80 percent and 20 percent, respectively). In short: the *number of fixations* radically increased owing to the instruction to search for possible friends or acquaintances in the film.

Furthermore, not only the number of fixations, but also the time spent looking at faces increased in the "acquaintance or friend searching" condition as compared to the "mere traveling" situation. This is illustrated in Figure 4, which presents data on *dwell time* on faces versus non-faces for all subjects again in proportions.

As Figure 4 shows, we detected a characteristic increase in time when subjects were watching faces in the "friend searching" condition as compared to the "mere traveling" condition. While in the unmarked, and possibly more natural, case of "mere traveling", fixation time on face and on non-face was almost equal (49 and 51 percentages), the task to search for a friend or acquaintance changed the amount of viewing faces versus non-faces dramatically. Under this second condition, the time spent gazing on faces increased to up to 82 percent, while watching non-faces dropped to 18 percent of all viewing time.

Our supposition during creating and conducting the visual interviews was that individual identification of people is produced by looking at other people's faces. This supposition was apparently confirmed. In a sense, the results described above are not quite surprising; indeed, they are replications of a stable and recurring result of eye-

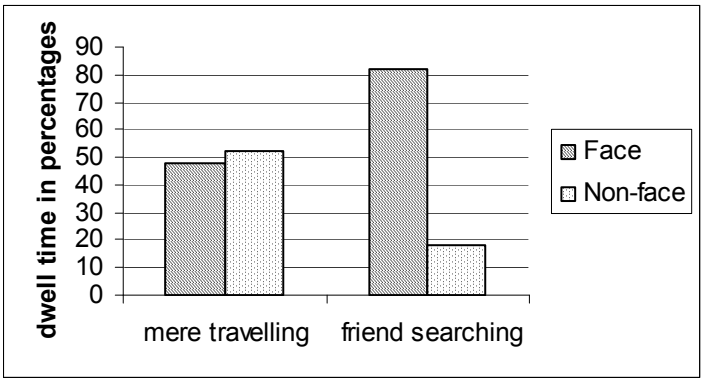

Figure 4. Proportion of dwell time on faces versus non-faces in "mere traveling" and "acquaintance or friend searching" conditions.

movement research since the work of Yarbus (1967). As he put it, the configuration of searching and examining pictures "is dependent not only on what is shown in the picture, but also on the problem facing the observer and the information that he hopes to gain from the picture" (Yarbus 1967: 194). The replication of these characteristics of gazing, however, indicates that our procedure of obtaining gaze pattern and the applied coding procedure were reliable.

Now, let us consider the pattern of gazing at overcrowded scenes in city dwellers versus country people. Although we consider the difference in gaze pattern between the two interview conditions an important finding, we were even more interested in the pattern of the more neutral – or at least less directly instructed – "mere traveling" condition. When in a crowded place, do people living in big cities use different gaze behavior as compared to people living in little towns or villages? Is their behavior in public places tangibly different with regards to viewing other people's faces? Or, to put it more theoretically, do they apply different gazing varieties, or are their habitual ways of watching, or – borrowing Bourdieau's (1977) term – the "habitus" of watching different in this regard? If country people try to abide strictly to the common rule of public behavior saying "greet all acquaintances and friends passing by", which city people do not follow for some (ecological and/or cognitive?) reasons, or follow more leniently, then a difference can be expected between the two groups, especially when they are simply in the "mere traveling" condition in a public place. If city people and country people exercised different gazing patterns indeed while watching the same scenes, in our view, this would be compelling evidence for different styles of gazing. Notably, if city people devoted fewer fixations to other people's faces and spent less time looking at them as compared to country people, then this finding could be attributed to a different gazing style. In this case, the answer to the above questions would be "yes". In short, we would gain evidence for the existence of different gazing styles.

Now, let us have a look at the data we obtained from the visual interviews conducted with the persons from the two groups. First, consider the distribution of the *number of fixations*.

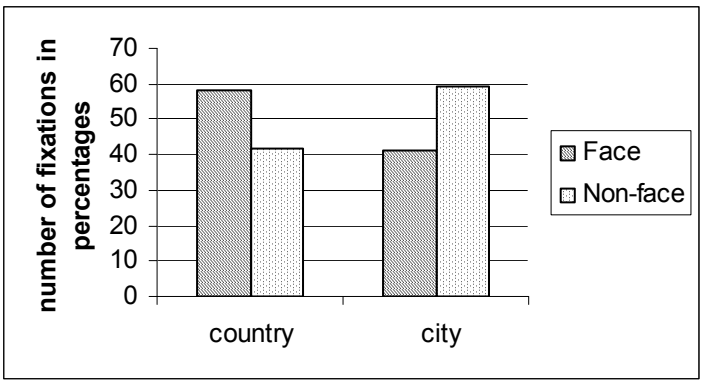

Figure 5. Proportion of number of fixations on faces in "mere traveling" condition for city group and country group.

People in the country group fixated on faces on considerably more occasions (58%) than on other aspects of the films (42%). At the same time, city people attested an opposite pattern. They fixated more on non-faces (58%) than on faces (42%). Figure 5 presents percentages of the number of fixations on faces for the city and the country group respectively.

The proportions of the number of fixations are informative with regard to the visual search strategy, and indicate a person's area of interest, or the most informative area for a person. More fixations on faces show a greater interest in other people's faces. In short, gazing at others' faces can very well be a way or method for realizing a search for possible acquaintances or friends in the mass of passers-by. Visual searching is a fast procedure, involving short fixations on the relevant areas of the image, thus a relatively great number of fixations indicates a searching behavior or method. On the other hand, it is not necessarily accompanied by increased dwell time. In any case, beside the number of fixations, the overall time spent on looking at faces would be a further constituent of a particular gazing style. Figure 6 presents our findings in this regard.

As Figure 6 shows, *dwell time* on faces was also different for the two groups in the predicted direction, although much less than in the case of the number of fixations. Overall, country people spent 52 percent of their fixation time on looking at faces, while 48 percent on looking at non-faces. At the same time, dwell time on faces was45 percent of their total fixation time while the remaining 55 percent was spent on non-face. It seems that country people gaze somewhat more often at faces than city people.

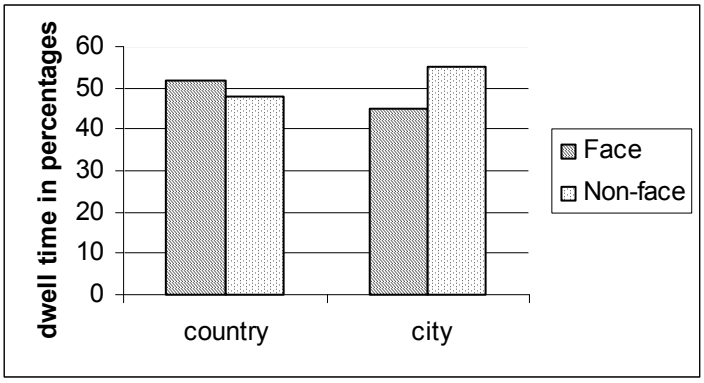

Figure 6. Proportion of dwell time on faces in "mere traveling" condition for city group and country group.

Collecting and analyzing more data will help us to decide whether this tendency, which is in accord with our expectations, can be taken as a solid and reliable pattern.

Our preliminary results show that, while in a public place, city people are less eager to watch other people's faces; they fixate less often; and spend somewhat less time fixating on faces. At the same time, country people fixate more often on faces and spend a little more time looking at them. It seems to us that the visual interview procedure we described above is a promising device for analyzing visual conduct, and eventually it could help us to decide whether our conjecture about the existence of two different gazing patterns according to place of residence is valid or not.

3. Discussion

A widely accepted model of cognition in general – and of visual search in particular – differentiates between "bottom-up" and "top-down" processes (Henderson 2007). Roughly speaking, "bottom-up" approaches highlight the role of environment in controlling gaze, particularly the role of such characteristics of image as intensity, contrast, edges, orientation, motion and color. On the other hand, "top-down" approaches emphasize such characteristics as knowledge structure, short and long term memory, and the goal of the person.

The effect of different instructions (in our case, the "mere travelling" versus "friend searching" conditions) on gazing pattern can be accounted by "top-down" processes. Furthermore, during our visual interviews, both groups attended the same film. Therefore, stimulus factors or "bottom-up" processes in themselves cannot account for the difference between the gazing of city versus country people. Consequently, in this case, some cogni-

tive properties of the persons, i.e. "top-down" processes, are to be relied on in accounting for the results. Note however, that within this framework, different "top-down" processes have to be postulated for people living in big cities and for persons living elsewhere. Apparently, this fact is not easily reconciled with a strictly individualistic psychology.

Some psychologists working with the eyetracker technique argue (Locher 1996) that a clear cut division between "cognitively-driven" and "sensory-driven" processes is not a really fruitful way of analyzing eye movements and gazing. Supposing an ongoing sequence of interaction between image and cognitive processes, Locher argues that a more circular account suits the case better. A rigid outside–inside attribution of causality is not very helpful in understanding a dynamic process. To our knowledge, the best available concept within cognitive psychology is still the concept of *schema*, as devised and characterized by Neisser some thirty years ago in his seminal work *Cognition and Reality* (Neisser 1976). In his view, a schema is an organized set of cognitive representations originated from earlier experiences. Its working is cyclical. It has three major aspects or phases: the first is an anticipatory process, which is based on previous experience and momentary goals, and actively prepares the mind for taking up information; the second is picking up the relevant information from the environment according to the anticipation; and, finally, the schema is modified by the new information. Thus, both the active role of the person, together with her/his knowledge structures and goals, and the environmental information, such as affordances, stimuli of his/her surroundings etc., are taken into consideration in the process during which the cognizer (the viewer) continuously accommodates himself/herself to the cognizant (image).

Indeed, this is a dynamic account, which postulates a continuous sequence of transactions between the mind and its environment, instead of separating them from each other. In the case under consideration, however, even within such a dynamic framework, there needs to be a specific reason to account for the similar schema in the two groups of different place of residence. After all, our demonstration indicates a similar gazing pattern for both groups on the basis of their place of residence, which is partly individual, since only persons can reside in a literal sense; but at the same time, it attests something "supra-individual", since a place of residence is not an individual's property. Thus, in psychological terms, if our preliminary results indicate some stable features, they can be accounted for by similarities in schema. However, even in this case, similarities as such are to be explained somehow.

Alternatively, or perhaps in parallel to referring to schemas, one can argue that there is something like a "viewing dialect" insofar as varieties of viewing are comparable to local language varieties – a concept stemming from sociolinguistics – or "viewing habitus" as "the durable installed generative principle of regulated improvisations" (Bourdieu 1977: 78) – a concept originating from sociology. In this way, a process evidently performed by an individual could be accounted for by certain social norms of conducting gaze which are characteristic to certain local communities.

4. Concluding remarks

As a summary of our study, we would like to formulate four related points. First, it seems to us that the above described visual interview method proved to be a useful and promising procedure in studying visual conduct.

Second, on a more theoretical level, our analysis indicates that gazing may be considered *as a habit*, which adapts to the actual visual environment in a flexible manner. Adaptation is influenced by the person's actual interest and could be modified by time at the initiation of the actor or other guiding agents. These agents that shape gazing conduct can be either persons, especially parents, who guide and guard the participation in children's interactions in private and public places, or new environmental, ecological arrangements, such as moving into a city from the country. In our view, it is meaningful to talk about habits of eyes as active viewing, and according to our theoretical proposal, these habits are at least partially socially established and maintained. Considering viewing as a habit opens up visual research to a more comprehensive interpretation of viewing, which includes its social determinants beyond the more usual environmental and individual factors.

Third, we proposed and exemplified that visual aspects of the organization of interaction may be investigated by the visual interview method. Given that the first move of an encounter is a summon, frequently performed by a greeting gesture, we identified individual recognition as "move number zero", and active visual scanning as "move number minus one" with regard to an emerging encounter or face engagement. Further, we described a preliminary study, which can be taken as a first step toward demonstrating that active visual scanning in public places is realized in variable gazing patterns according to place of residence, or *visual dialects*. Insofar "move number minus one" of an encounter is performed differently by city people and country people, i.e. country people are more active in watching other people's faces than city people, these varieties of gazing cannot be accounted for by strictly individualistic psychological mechanisms.

Finally, we would like to underline the fact that our study is an attempt at demonstrating empirically the feasibility of the conceptual insights of some founding fathers of sociology, like Tönnies (1887/2001) and Simmel (1903/1950). In this context, collecting and analyzing further empirical data on the gazing conduct of country people and city people could contribute to a perennial topic of sociology, i.e. the oft-detailed difference in life-style, interpersonal relationships and mentality among city dwellers versus country people could be given empirical support or refutation – at least at the level of visual interaction. Our preliminary findings indicate that, in terms of gazing, there may be a difference between the ways or methods people relate to each other in cities and in the country, which is a major precondition of establishing various types of face to face interactions or encounters. Although these social differences seem to be realized individually by our eyes, they may very well indicate difference in "habitus".

REFERENCES

Bourdieu, P. 1977. *Outline of a theory of practice*. London: Cambridge University Press.
Bridgeman, B., A.H.C. Van der Hejiden and B.M. Velichkovsky. 1994. "A theory of visual stability across saccadic eye movements". *Behavioral and Brain Sciences* 17(2). 247–292.
Böhme, M., A. Meyer, T. Martinetz and E. Barth. 2006. "Remote eye tracking: State of the art and directions for future development". (Paper presented at the 2nd Conference on Communication by Gaze Interaction – COGAIN 2006, September 4–5, Turin, Italy.)
Goffman, E. 1963. *Behavior in public places*. New York: The Free Press.
Goffman, E. 1983. "The interaction order: American Sociological Association, 1982 Presidential Address". *American Sociological Review* 48(1). 1–17.
Goodwin, C. 2001. "Practices of seeing visual analysis: an ethnomethodological approach". In Van Leeuwen, T. and C. Jewett (eds.), *Handbook of visual analysis*. London: Sage. 157–183.
Harré, R. and G. Gillett. 1994. *The discursive mind*. London: Sage.
Heath, C. 1997. "The analysis of activities in face to face interaction using video". In Silverman, D. (ed.), *Qualitative research: Theory, method and practice*. London: Sage. 183–201.
Henderson, J.M. 2007. "Regarding scenes". *Current Directions in Psychological Sciences* 16(4). 219–222.
Locher, P.J. 1996. "The contribution of eye-movement research to an understanding of the nature of pictorial balance perception: A review of the literature". *Empirical Studies of the Arts* 14(2). 143–163.
Miall, R.C. and J. Tchalenko. 2001. "A painter's eye movements: A study of eye and hand movement during portrait drawing". *Leonardo* 34(1). 35–40.
Mogg, K., B.P. Bradley, M. Field and J. De Houwer. 2003. "Eye movements to smoking-related pictures in smokers: Relationship between attentional biases and implicit and explicit measures of stimulus valence". *Addiction* 98(6). 825–836.
Neisser, U. 1976. *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: W.H. Freeman.
Psathas, G. 1990. "Introduction: Methodological issues and recent developments in the study of naturally occurring interaction". In Psathas, G. (ed.), *Interaction competence*. London: University Press of America. 1–31.
Rayner, K. 1998. "Eye movements in reading and information processing: 20 years of research". *Psychological Bulletin* 124. 372–422.
Riley, P. 2007. *Language, identity and culture*. London: Continuum.
Sekuler, R. and R. Blake. 1994. *Perception*. New York: McGraw-Hill.
Simmel, G. [1903] 1950. "The metropolis and mental life". In Wolff, K. (ed.), *The sociology of Georg Simmel*. New York: Free Press. 409–424.
Tanenhaus, M.K., J.S. Magnuson, D. Dahan and C. Chambers. 2000. "Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing". *Journal of Psycholinguistic Research* 29. 557–580.
Tönnies, F. [1887] 2001. *Community and society*. Cambridge: Cambridge University Press.
Yarbus, A. 1967. *Eye movements and vision*. New York: Plenum.
Unema, P., S. Pannasch, M. Joos and B.M. Velichkovsky. 2005. "Time-course of information processing during scene perception: The relationship between saccade amplitude and fixation duration". *Visual Cognition* 12(3). 473–494.
Velichkovsky, B.M., A. Rothert, M. Kopf, S.M. Dornhofer and M. Joos. 2002. "Towards an express-diagnostics for level of processing and hazard perception". *Transportation Research Part F: Traffic Psychology and Behaviour* 5(2). 145–156.

Woodworth, R.S. and H. Schlossberg. 1963. *Experimental psychology*. New York: Holt, Reinhart and Winston.

APPENDIX

Summary data on the participants of the visual interviews.

City dwellers

Gender	male	female	female	female	male
Age	38	33	28	41	30
Education	University	University	High School	University	University
Present place of residence	city	city, since 1994	city, since 1992	city	city, since 1991
Previous place of residence	–	little town	little town	–	little town
Visiting big cities	–	–	–	–	–

Country dwellers

Gender	female	male	male	male	female
Age	33	38	25	18	18
Education	High School	Vocational School	Vocational School	High School	High School
Present place of residence	village	village	little town	little town	little town
Previous place of residence	Attending school in Budapest 1985–1989	Attending school in Budapest 1989–1993	Attending school in Budapest 1998–2000	–	–
Visiting big cities	Once a month	Once a month	Once a month	A few times a year	A few times a year