

e t c



*Dedicated to quantitative empirical
studies of culture*

*RAM Verlag
2010*

Empirical Text and Culture Research

4

ETC

Empirical Text and Culture Research

4

*Dedicated to quantitative empirical
studies of culture*

**RAM-Verlag
2010**

Editor:

Andrew Wilson (Lancaster University, UK)

Editorial board:

Valery Belyanin (CERES, University of Toronto, Canada)

Katerina Frantzi (University of the Aegean, Greece)

Robert Hogenraad (Catholic University of Louvain, Belgium)

Dean McKenzie (Monash University, Australia)

Josef Schmied (Chemnitz University of Technology, Germany)

Kaoru Takahashi (Toyota National College of Technology, Japan)

Editorial contact address

Dr. Andrew Wilson
Department of Linguistics and English Language
County South
Lancaster University
Lancaster LA1 4YL, UK

Email: eiaaw@exchange.lancs.ac.uk

Fax: +44 1524 843085

Orders for CD-ROM's or printed copies to

RAM-Verlag RAM-Verlag@t-online.de

Downloading: <http://www.ram-verlag.de>

Empirical Text and Culture Research

Scope

The purpose of ETC is to publish culturally oriented research that is both empirical and quantitative in its approach.

'Culture' is understood by ETC to encompass both 'high' culture (such as literature, art, and music) and the many aspects of everyday life and culture (such as fashion or cuisine). The 'world views' of different cultures and subcultures, and the relationship between language and culture, will also be key issues for the journal.

Whilst work in the framework of systems-theoretic constructivist research is especially welcome, ETC aims to publish all forms of empirical quantitative research that fit its parameters – for example, corpus-based research; computer-assisted content analyses; studies using the semantic differential, sentence completion, and word association techniques, and so on.

ETC will normally not publish papers of a purely theoretical or qualitative nature, unless their implications for quantitative empirical research are very clear.

We accept both full-length papers (approximately 10 to 20 double-spaced pages) and short notes (of around 5 pages or less).

Reviews of relevant books, software, conferences, etc., as well as historical and biographical studies related to empirical cultural research, are also welcome.

If anyone is unsure about the suitability of their topic or approach for ETC, they are invited to send a short structured abstract to the editor in advance of a formal submission.

From time to time, the editorial board may actively commission articles or reviews.

Language policy

For the journal to have the maximum international impact, only papers written in English will be published in ETC.

Peer review

All papers will be peer reviewed by the editor, another member of the editorial board, or some other qualified person appointed by them. We will aim to move swiftly and normally provide a response within one month of submission.

Contents

Francesca Bianchi

*Understanding culture. Automatic semantic analysis
of a general Web corpus and a corpus of elicited data* 1-29

Arjuna Tuzzi, Ioan-Iovitz Popescu, Gabriel Altmann

The golden section in texts 30-41

Bernadette Péley, János László

Character functions as indicators of self states in life stories 42-49

Katalin Szalai, János László

*Activity as a linguistic marker of agency: Measuring in-group versus
out-group activity in Hungarian historical narratives* 50-58

Réka Ferenczhalmy, János László

In-group versus out-group intentionality as indicators of national identity 59-69

Enikő Gyöngyösiné Kiss

Personality and the familial unconscious in Szondi's fate-analysis 70-80

Bruno Gonçalves, Ana Ferreira, Mátyás Káplár, Enikő Gyöngyösiné Kiss

*Comparing the Szondi Test results of Hungarian and Portuguese
community samples* 81-89

Anita Deák, Laura Csenki, György Révész

*Hungarian ratings for the International Affective Picture System (IAPS):
A cross-cultural comparison* 90-101

Attila Oláh, Henriett Nagy, Kinga G. Tóth

Life expectancy and psychological immune competence in different cultures 102-108

Paszkál Kiss

*Hungarian answers to an international crises: Parliamentary debates
during the 1999 Kosovar conflict* 109-118

Kaoru Takahashi

A Study of Sociolinguistic Variables in the British National Corpus 119-134

Andrew Wilson

Fetishism and anxiety: A test of some psychodynamic hypotheses 135-143

Understanding culture: Automatic semantic analysis of a general Web corpus and a corpus of elicited data¹

Francesca Bianchi²

University of Salento, Italy and Lancaster University, GB
Dip. Lingue e Letterature Straniere, Via Taranto 35, 73100, Lecce, Italy

Abstract

This study investigated the suitability of different methodological approaches to automatic semantic tagging in the analysis of cultural traits as they emerge from subjective meaning reactions to given words (EMUs). Elicited data from British native speakers were collected and coded manually and with an automatic tagging system (Wmatrix). The results of manual coding were then compared to the results produced by Wmatrix, at different levels and using a variety of methods. Subsequently, automatic tagging was applied to 10,000 sentences containing the node word extracted from a general Web corpus, and the results of the Web corpus were compared to those of the elicited data. Though further investigation is needed, each of the experiments described provide results that are relevant for the definition of a method in the use of large corpora for the extraction of EMUs.

Keywords: culture, corpora, semantic tagging, Wmatrix, methodological issues

Introduction

This paper explores some methodological approaches to semantic analysis applied to the understanding of the cultural traits of a given group (the English) as they emerge from subjective meaning reactions elicited by a particular word. This is a first step in a wider project aimed at applying some of the analytical resources of corpus linguistics to cultural studies and marketing research. Cultural studies and marketing research are two wide areas which include several sub-areas, each one characterised by many different theoretical and analytical standpoints. The following paragraphs will briefly outline the theoretical frameworks that inspired the current investigation.

¹ This article is a reviewed and expanded version of a paper of the same title presented at the Corpus Linguistics conference, Liverpool, 20-23 July 2009.

² Author's e-mail: francesca.bianchi@unisalento.it

Culture, language, and semantics

How to define culture has long been a debated issue in diverse disciplines, such as literature, art, archaeology, philosophy, anthropology, semiotics, and more recently linguistics, translation studies, and marketing. Despite the different conceptualization of culture within each scientific discipline, and the particularities of individual theories, some common ideas seem to be shared by the scientific community at large. According to these, one may say that culture is a highly complex and variegated social and semiotic event which is acquired through processes of transmission of information; it may develop both diachronically and synchronically; and some cultural aspects may be visible in everyday life through language or other man-made products. However, there are two aspects of culture which are of particular relevance here: it emerges through language, and is shared to a high degree by members of the same community.

The existence of a link between language and culture has long been postulated across disciplines (see for example: Humboldt, 1836; Malinowski, 1923; Sapir, 1949, and Whorf, 1965, in social anthropology; Lotman, 1993, in semiotics; Halliday, 1978, and Fairclough, 2003, in linguistics; Cavalli-Sforza, 1996, in genetics). Indeed, language is the primary means by which we describe the world and express our thoughts and ideas. Beliefs and judgements, along with values and value orientations, pertain to what Hall (1989) calls ‘the informal level of culture’. This level of culture – which is distinct from other two levels: the technical and the formal ones – can neither be taught nor learnt; it is passed on and acquired unconsciously, or ‘out-of-awareness’. It is at this informal level where people normally react in everyday life and communication. Beliefs, values and value orientations permeate our thoughts and govern our mental and physical actions, even if we do not realise it. Thus, they also permeate our language, at all possible levels: from semantics to grammar, from pragmatics to discourse structure.

This work shall consider the level of semantics. Inspiration was taken from the concept of Elementary Meaning Units (or EMUs), a concept that was first defined by Greenberg (1960, cited in Szalay & Maday, 1973, p. 34), and subsequently used in anthropology for the purpose of cross-cultural comparisons. Elementary Meaning Units are subjective meaning reactions to individual words. They include three major dimensions which are all equally important in the study of culture, namely composition, dominance, and organisation. Indeed, psychological meaning depends on the composition of several distinct elements, including visual imagery, context of use, and affective reactions. Dominance, on the other hand, is measured in terms of frequency: some EMUs are dominant in a given cultural group, as they are more frequent than any others and their higher frequency influences cognitive processes. Finally, the psychological lexicon is organised according to networks based on affinity between elements. In this respect, EMUs cluster into semantic domains, or “large units of cognitive organisation” (Szalay & Maday, 1973, p. 34).

EMUs have traditionally been analysed using elicited data. Szalay and Maday (1973) used verbal associations and semantic grouping of verbal responses in the native language in question to assess subjective, or implicit, culture, i.e. “psychological variables, images, attitudes, and value orientations” (Szalay & Maday, 1973, p. 33) of American and Korean subjects. They analysed EMUs in three semantic domains: *education*, *manners* and *family*. Their method was based on eliciting verbal associations using lists of words provided by the participants in the experiment; responses to the words in the first list were then compared, and those in common were used as stimulus words for the next verbal association task, and so on. Subsequently, the responses were grouped into fewer categories and analysed in order to establish the affinity structure and

cognitive organisation of each domain word. This way the authors managed to establish that the American and Korean cultures diverge in their views of *family* and *education*. In fact while EMUs *polite*, *greeting*, *manners*, and *to bow* consistently showed low affinity with the domains under consideration in the data from American students, they showed high affinity in those from Korean students. The authors concluded that free verbal associations obtained in the native language, solicited with the method described, “provide empirical data on the denotative as well as the connotative components of meaning” and “allow to reconstruct culture-specific cognitive organisation by its main dimensions” (Szalay & Maday, 1973, p. 41).

A clear parallelism can be seen between EMUs and empirical collocations (see Evert, 2008, for a detailed discussion of the concept of collocation in corpus linguistics).³ Empirical collocations are words that co-occur in the same textual environment; frequency of co-occurrence determines collocational strength. The collocates of a node word, once grouped into semantic fields or domains, show its semantic preference (Partington, 2004). Analogously, EMUs co-occur in the same psychological environment as the word that triggers them, and they all show high collocational strength to the node word. Higher frequency of one EMU over another could thus be an indication of a cultural (vs. an individual) origin of the EMU itself.

This last observation may be better understood considering Fleischer’s theory of culture (Fleischer, 1998). Fleischer sees culture as a stratified system of several interacting and at times (partially or totally) overlapping social systems and sub-systems, each expressing itself in a different discourse system. Discourse, being a semiotic element, is composed of symbols. Symbols are made up of three elements, which Fleischer calls ‘core’, ‘current field’, and ‘connotational field’. Both core and current field are expressions of the cultural system under analysis. The core is a stable semantic element, while the current field is a rather generalised, though not yet stabilised, element. Finally, the connotational field is an expression of individual meaning. These elements allow us to interpret the level of rooting of a word in the given culture: if the connotational field (individual meanings) predominates in the understanding of a word, then that word is not a symbol. When current field predominates, the given word is in the process of becoming a symbol. Finally, when core meanings predominate, we are dealing with a very strong type of symbol. Therefore, high-frequency EMUs correspond to the core and current elements of the symbol under analysis and are deeply rooted in the culture, while low-frequency EMUs belong to the connotational field and are connected to the individual, not the community.

Fleischer’s theory of collective symbols has been empirically tested in a few studies, including Fleischer (2002). In order to identify the semantic profiles and level of conventionalisation of the image of drinks in Poland, France and Germany (and their corresponding cultures) Fleischer (2002) presented groups of native volunteers with a list of drinks and beverages and asked them to write whatever came to their minds for each of the names in the list. The respondents’ answers were grouped under three broad semantic categories: characteristics and connotations (*Konnotationen & Images*), trademarks and proper names (*Umschreibungen & Marken*), and evaluations (*Wertungen*). Within each category, words and phrases were then grouped into unlabelled semantic sub-categories. Hapax legomena were considered to be connected to individual feelings and preferences and were thus ignored. Looking at the ratio between the described

³ Interestingly, some recent empirical research has shown “a direct predictive relationship between the statistics of word co-occurrence in text and the neural activation associated with thinking about word meanings” (Mitchell, Shinkareva, Carlson, Chang, Malave, Mason, & Just, 2008: 1191; Murphy, Baroni, & Poesio, 2009). These results suggest that a direct relation between co-occurrence of words in text and the mental lexicon may exist, though further research is needed in this field.

features (types) and individual instances of each feature (tokens), and confidence intervals based on means and standard deviations, three levels of conventionalisation were established. Results showed differences between the three cultures, in terms of semantic profile and level of conventionalisation. These differences were then discussed with reference to historical, sociological and more generally cultural events.

Analytical semantic descriptions of the meaning components (be they cultural or individual) of a word or concept like the ones provided by Fleischer, and Szalay and Maday and discussed above could, in my opinion, be of great help in marketing research.

Marketing research

Quoting from Hair, Bush, and Ortinau (2009, p. 4), marketing research is “the function that links an organisation to its market through the gathering of information”. This is a broad definition that encompasses several types of data gathering and analytical activities aimed at providing decision makers with information that might help them plan future action and interaction with the desired audience.⁴

Data collection in marketing research is carried out on two separate types of sources. Primary information – i.e. “information specifically collected for a current research problem or opportunity” (Hair *et al.*, 2009, p. 37) – is gathered through qualitative research questioning techniques, such as in-depth interviews and focus groups, or through quantitative techniques, which include case studies, as well as various types of interviews and tests. These techniques collect elicited data but are highly expensive and time consuming tasks. Secondary information – i.e. “information previously collected for some other problem or issue” (Hair *et al.*, 2009, p. 37) – may be customer-volunteered information in the form of “information gathered from electronic customer councils, customer usability labs, e-mail comments, chat sessions, and so forth”, but also “data collected by the individual company for accounting purposes or marketing activity reports”, or data collected by outside agencies, associations or periodicals (Hair *et al.*, 2009, pp. 114-115). Growing emphasis has recently been put on secondary data, partly as a consequence of the development of the Internet (Hair *et al.*, 2009, p. 37), and Internet work seems to be gradually replacing field work.

In the current work, elicited data gathered through sentence-completion and sentence-writing tests will be used as a starting point. To these, analytical methods typical of corpus linguistics will be applied. Indeed, if automatic analytical tools of the types used in corpus linguistics, and corpora of some sort gave the same results as more traditional marketing research techniques, marketing research could benefit from a wider range of fast and inexpensive methods.

Methodological ideas from corpus linguistics

A number of interesting quantitative studies of culture-specific traits can also be found in corpus linguistics (see Bianchi, 2009). Among them, some have taken advantage of frequency word lists, subsequently grouped into semantic categories. Leech and Fallon (1992) used frequency tables to highlight cultural differences between the American and British cultures as they emerged from the

⁴ As such, marketing research is neither good nor bad in itself: it is simply gathering of information. The use that decision makers make of it, though, may be targeted to gaining personal advantage (as in private business advertisements) or to higher and ‘friendlier’ goals, as is the case with ethical and social advertising campaigns. This note is in reply to some criticism I have recently received for associating myself with marketing research.

Brown and LOB corpora. The words in the frequency tables were grouped into semantic categories, and categories where frequency differences were noticeable were identified. The authors then used quantitative differences to draw generalised conclusions about the two cultures. This study was duplicated by Oakes (2003) using the FROWN and FLOB corpora. In both studies the American culture emerged as masculine, militaristic and dynamic, driven by high ideals, technology, activity and enterprise; the British culture as “given to temporizing and talking, to benefiting from wealth rather than creating it, and to family and emotional life, less actuated by matters of substance than by considerations of outward status” (Leech & Fallon, 1992, pp. 44-45). Muntz (2001) applied Leech and Fallon’s methodology in order to highlight cultural differences between British English and Australian English. Furthermore, Schmid (2003) applied Leech and Fallon’s analytical method to the spoken part of the BNC. However, as this author was looking for confirmation to Deborah Tannen’s theory on gender differences, only the words from 18 domains inspired by Tannen’s (1990) book were analysed. Finally, Bianchi (2007) used a large general corpus and a small-to-medium-sized specialised Web corpus to highlight EMUs to chocolate in contemporary Italian society. Concordances were generated for the Italian words for chocolate, and each concordance line was manually classified in terms of semantic fields, i.e. the main topic(s) mentioned in the text segment. Semantic fields were then grouped into higher-order semantic categories, and results from the two corpora were compared in order to highlight what could probably be considered long-existing and well-established EMUs for chocolate in Italian society.

Research in the field, therefore, suggests that corpora, including those compiled from the Web, can be a suitable data source for cultural analysis. Web corpora, in particular, could theoretically be an interesting choice. First of all, they are rather quick to assemble, compared to corpora from more traditional textual sources, given that suitable text can be downloaded automatically using spidering tools (e.g. see the procedure described in Baroni & Bernardini, 2004). Also, Web corpora tend to include up-to-date text and language, an important point considering that time span is a major issue in cultural studies. Indeed culture evolves over time, and EMUs may change very quickly, sometimes in less than a year (Nobis, 1998). Furthermore, recent research (Fletcher, 2004; Sharoff, 2006; Ueyama, 2006; Baroni & Ueyama, 2006) suggests that large general Web corpora assembled using spidering systems and following specifically reasoned basic criteria may be considered as representative as analogous manually-collected corpora, at least in terms of text type and domain coverage, if not of register (which is of no relevance in the analysis of EMUs).

When using Web corpora for cultural analysis, however, some caveats should also be considered. First, if the research target is a whole single culture, it seems important that the corpus include a wide variety of texts by different authors and that it cover a limited time span (Bianchi, 2007). Second, authorship is still an unresolved problem: neither the fact that a page is written in a specific language, nor that it is published in a specific country guarantees that the author is native to that language/culture. This is particularly true in the case of English, as it has been gradually establishing itself as a *lingua franca* and as ‘the’ language of the Internet (Crystal, 2003). For these reasons, a direct comparison with elicited data is necessary before Web corpora can be finally considered an adequate source of data for marketing research.

To sum up, two common points can be seen in marketing research methods and the cultural studies quoted in the previous sections: the use of elicited data; and analytical methods based on manual semantic coding. For this reason this study will use elicited data and manual coding as a primary source of information. The results of manual coding will then be compared to those obtained with an automated procedure, in order to investigate the possibilities offered by an

automatic on-line tool for semantic tagging for the purpose at hand. The semantic tagger under analysis – the USAS tagset in the Wmatrix interface, which will be described in the following section – was originally developed for automatic content analysis of elicited data (Wilson & Rayson, 1993) and has been used with interesting results in diverse corpus linguistic studies on a range of different topics, from stylistic analysis of prose literature to the analysis of doctor-patient interaction, and from translation to cross-cultural comparisons (see <http://ucrel.lancs.ac.uk/usas>). One of the aims in this study was to see whether and how this semantic tagger could be used for the extraction of EMUs and how it compared to manual coding. Furthermore, elicited data were compared to a large Web corpus for general purposes, following the assumption that if results from Web corpus data matched results from surveys, marketing research could benefit from a quick and low cost analytical method.

Material and method

For the current study two node words were selected – *chocolate* and *wine* – and each of them was analysed in two different sets of data: a corpus of elicited data; and a large Web corpus for general purposes. These data sources are described in the following paragraphs.

The elicited data were collected specifically for the purpose of this study. The data were elicited by means of questionnaires with sentence completion and sentence writing tasks. In fact, the questionnaires, which featured a picture illustrating the node word, began with the following six completion sentences:

- | | |
|---|---|
| 1. Whenever I think of chocolate I | / Whenever I think of wine I |
| 2. Chocolate reminds me of | / Wine reminds me of |
| 3. The picture on the top leads me to | |
| 4. Chocolate can | / Wine can |
| 5. I would use chocolate to | / I would use wine to |
| 6. It's common knowledge that chocolate | / It's common knowledge that wine |

This task was followed by a request to write 20 sentences using the node word given.

The questionnaires were first circulated via e-mail, then distributed manually within the University campus at Lancaster. This allowed us to reach a total of about 90 English native speakers aged 18 to 60. However, two thirds of the respondents were university students in the 18-25 age group. All respondents completed the first task, while in the sentence-writing task some wrote less than 20 sentences, or even no sentence at all. Using the data thus gathered, two elicited corpora were created, as detailed in Table 1. As the first task in each questionnaire (6 items) was a sentence completion exercise, the corpora were saved in two different formats: Format 1 (F1) which included the words given in the first six sentences; and Format 2 (F2), which did not include the given text. F1 was used when performing manual coding of the elicited data; F2 when performing automatic tagging.

Table 1. Elicited data summary.

	Chocolate	Wine
Total n. of respondents	87	91
Total n. of sentences	1888	1938
Mean n. of sentences (SD)	21.7 (SD = 6.58)	21.3 (SD = 6.57)
Total n. of words (F1)	12946	13740
Running words in tagged corpus (F2)	9967	10967

A few of the sentences in the elicited data (15 for *chocolate* and 21 for *wine*) were connected to the questionnaire or the situation, rather than to the node word (e.g.: *Sorry I have revision to do; I feel daft writing about chocolate; I don't know as much about chocolate as I do about wine*), or were ambiguous in their reference to the node word or pertinence to the purpose of the survey (e.g.: *Wine begins with w; There is no wine in winegums*), but it was decided not to remove them from the elicited corpora. In fact, deleting sentences of this type from the elicited data, but not from the Web corpora would have been pointless, if not altogether methodologically wrong. At the same time it would be impossible to identify (and remove) 'irrelevant' sentences from the Web corpus, given its size and the fact that in some cases the pragmatic context of the original texts might be unintelligible.

As source of Web data, the UKWAC corpus (Baroni & Kilgarrieff, 2006; Baroni, Bernardini, Ferraresi, & Zanchetta, 2008) was used. This is a large general corpus of about two billion running words, created from the Web using spidering tools; the corpus is lemmatised and POS tagged with Tree-Tagger. UKWAC was accessed using the Sketch Engine (www.sketchengine.co.uk; Kilgarrieff, Rychly, Smrz, & Tugwell, 2004), an on-line interface that allows concordancing and other linguistic queries. The interface was set to save 10,000 full sentences, which led to the creation of two sub-corpora: the chocolate sub-corpus, with 9944 sentences and 407962 running words; and the wine sub-corpus, with 9960 sentence and 349740 words.

The elicited data underwent both manual coding and automatic tagging. The Web data, on the other hand, were only tagged automatically. The following sections describe the coding and tagging processes.

Manual coding

The manual coding task was performed following the steps suggested by Neuedorf (2002). These include the creation of an initial Codebook, followed by several cyclical phases of coder training, coding and discussion, followed by codebook revision.

Before starting the coding process of the elicited data, a Codebook was drafted which includes a detailed description of the coding scheme (with examples) and of its origin, and instructions on how to apply the coding scheme in the task at hand. The coding scheme described in the Codebook originates in a preliminary experiment of manual coding of Web data focusing on the node word *chocolate* in Italian (Bianchi, 2007) and in English (unpublished). The original codes were applied, discussed and reviewed twice before being included in the Codebook, version 1. The two sets of elicited data – F1 Chocolate and F1 Wine – were then coded by two separate coders who had received specific training on the use of the coding scheme. During the coding procedure the two coders met twice to discuss the need for further fields and/or areas. When a new semantic field was agreed upon and added to the list, each coder reviewed the sentences s/he had already tagged. Thus, the coding scheme grew from 15 Conceptual domains and 83 Semantic fields to the 15 Conceptual domains and 101 Semantic fields listed in Table 2 (Codebook, version 2).⁵ Semantic fields and conceptual domains represent two hierarchical levels of semantic analysis, the latter including superordinate, broader categories.

⁵ The coding scheme includes also assessment of semantic prosody, but this is not illustrated here, as it was not considered in the current analyses.

Table 2. Coding scheme.

Conceptual domains	Semantic fields
Food	Product/shape; Bakery/cooking; Manufacturing; Food; Composition; Recipe; Drink; Storage; Serving
Health & Body	Health; Medicine; Body; Beauty
Events	Language/etymology; Economy; Religion/mythology; War; History; Law; Event; Transaction; Fair trade; Time; Work; Driving; Excessive drinking; Holidays
Feelings & Emotions	Senses; Love; Desire; Pleasure; Sex; Happiness; Seduction; Mood; Passion; Competitiveness; Memory; Surprise; Loneliness; Freedom; Persuasion; Guilt; Comfort; Relax; Peace; Bribing; Confidence
People	Women; Men; Gay; Children; Posh; Friendship; Royalty; Sharing/society; People; Family; Age
Geography	Geographical locations; Spreading
Imagination	Fantasy/magic; Dream
Loss & Damage	Theft; Drugs and addiction; Hiding
Ceremonies	Ceremonies; Party; Gift
Environment & Reality	Nature; Animals; House; Dirt; Technology
Culture	Artistic production; Culture; Studying/intellect
Life	Future; Existence
Features	Quality/type; Colour; Sweet; Genuineness; Energy; Taste/smell; Quantities; Price; Packaging; Physical properties
Sports	Sports
Comparison	Comparison

In the manual coding task, the unit of data collection was the questionnaire, while the unit of analysis was the sentence. Coding was done manually and required the coders to assign one or more semantic fields (chosen among the ones given) to whole sentences on the basis of their assessment of the semantic fields that were explicitly or implicitly mentioned in the given sentence.

Decisions were usually triggered by specific words in the sentence (e.g. *Very good chocolate may be expensive* = PRICE; *Chocolate is good for your health* = HEALTH), but also by context (e.g. *So is Bulgarian wine* can only be understood in connection to the sentence that precedes it: *Chilean wine is good*), and/or general knowledge of the world (e.g. *I eat chocolate before sitting an exam* = ENERGY, because it's common knowledge that an exam is a hard task that drains your energies). In cases of disagreement between the two coders (about 3%), the suggestions of both were accepted.

Automatic semantic tagging

Automatic semantic tagging was applied to both the elicited data and the Web data using Wmatrix (Rayson, 2008), a fully-automated and user-friendly on-line interface. This system – developed at the Lancaster's University Centre for Computer Corpus Research on Language (UCREL) – works on any given text file in English.

In Wmatrix, semantic tagging is preceded by POS tagging and lemmatisation. POS tagging is performed using CLAWS - Constituent Likelihood Automatic Word-tagging System (Garside & Smith, 1997) and its standard CLAWS 7 tagset.⁶ This probabilistic tagger, developed at UCREL and used for tagging the BNC, reaches an accuracy of 96-98 % (Rayson, Archer, Piao, & McEnery, 2004). The semantic tagging component (described in Wilson & Rayson, 1993; Rayson *et al.*, 2004; and in Archer, Rayson, Piao, & McEnery, 2004) includes a single word lexicon of 42,000 entries, and multi-word expression (MWE) templates, with 18,400 entries in all. Furthermore, it includes context rules and disambiguation algorithms for the selection of the correct semantic category. This semantic tagging process performs with a 92% accuracy rate (Piao, Rayson, Archer, & McEnery, 2004, quoted in Archer *et al.* 2004).

The semantic categories used in the system were originally based on the *Longman Lexicon of Contemporary English* (LLOCE) (McArthur, 1981), though some changes were subsequently made (Rayson *et al.*, 2004). The current ontology includes 21 fields (Table 3), subdivided into 232 categories with up to three subdivisions, for a total of 453 tags.

Table 3. Semantic fields in the UCREL Semantic Analysis System tagset.

A - General & Abstract Terms	N - Numbers & Measurement
B - The Body & the Individual	O - Substances, Materials, Objects & Equipment
C - Arts & Crafts	P - Education
E - Emotional Actions, States & Processes	Q - Linguistic Actions, States & Processes
F - Food & Farming	S - Social Actions, States & Processes
G - Government & the Public Domain	T - Time
H - Architecture, Building, Houses & the Home	W - The World & Our Environment
I - Money & Commerce	X - Psychological Actions, States & Processes
K - Entertainment, Sports & Games	Y - Science & Technology
L - Life & Living Things	Z - Names & Grammatical Words
M - Movement, Location, Travel & Transport	

Clearly, this semantic structure is rather different from the one developed and used in the manual tagging process. However, as we shall see in the following paragraphs, comparisons are still possible, by applying a conversion process similar to that used for matching the UCREL semantic taxonomy to that of the Collins English Dictionary (CED) and described by Archer *et al.* (2004).

At the end of the tagging process, Wmatrix publishes the output in several different formats, including frequency word lists of the untagged, POS tagged, and semantically tagged versions of the file. Furthermore, it offers features for generating keyword lists, using the BNC as reference corpus.

Matching manual tags to automatic tags

To allow comparison, the USAS tags were matched to the semantic fields used in the manual coding of the elicited data. For each tag, matching was accomplished by looking at the prototypical examples provided in Archer, Wilson and Rayson (2002), imagining them in the given

⁶ List of tags available at: <http://ucrel.lancs.ac.uk/claws7tags.htm>.

context (i.e. next to the words *chocolate* and *wine*, but also in the wider context of general speech), and finding a suitable semantic field in the manual tagging list. Examples of matching are provided in Table 4.

Table 4. Conversion schemes: some examples.

USAS tag	USAS semantic category	Chocolate manual coding	Wine manual coding
O4.6+	Temperature: Hot / on fire	// Drink [Food] // Other	// Storage [Food] // Other
O1.1	Substances and materials: solid	// Food [Food] // Other	// Food [Food] // Other
I2.2	Business: Selling	Transaction [Events]	Transaction [Events]
X3.1	Sensory: Taste	Taste [Features]	Taste [Features]
E2-	Dislike	Passion [Feelings & Emotions]	Passion [Feelings & Emotions]
L1+	Alive	Existence [Life]	Existence [Life]
S3.1	Personal relationship: General	Friendship [People]	Friendship [People]
A2.1+	Change	Other	Other
A1.5.1	Using	Other	Other

In the table, in the manual coding columns, the first word or expression is the semantic field, while the one in square brackets is the corresponding conceptual domain. Double slashes (//) indicate that matching is ‘one to many’. The word ‘Other’ indicates no matching.

Different conversion schemes were necessary in order to account for the different fields of the two key words. For example, the elicited corpus showed that USAS tag O4.6+ (TEMPERATURE: HOT / ON FIRE), which corresponds primarily to the word *hot*, tends to refer to different semantic fields when next to the word *chocolate* or *wine*: if chocolate is hot, it is a drink; if wine is hot, we are talking about a storage issue. However, given that both chocolate and wine belong to the same general category of food and drinks, the two conversion schemes show a limited number of differences. A given USAS tag could match one or more categories of the manual codes, or even none of them. Matching was not sought for categories indicating logical or grammatical relations (Table 5). Indeed these categories were disregarded in all the analyses. This conversion scheme was used in Steps 1 and 2.

Table 5. Categories excluded from analysis.

code	description	code	Description	code	description
Z4	Discourse Bin	Z99	Unmatched	A13.3	Degree: Boosters
Z5	Grammatical bin	A7	Probability	A13.4	Degree: Approximators
Z6	Negative	A7+	Likely	A13.5	Degree: Compromisers
Z7	If	A7-	Unlikely	A13.6	Degree: Diminishers
Z7-	Unconditional	A13	Degree	A13.7	Degree: Minimisers
Z8	Pronouns	A13.1	Degree: Non-specific	A14	Exclusivisers/particularisers
Z9	Trash can	A13.2	Degree: Maximisers	N1	Numbers

One of the major issues in matching two different schemes of this type is how to distribute frequency in the case of ‘one-to-many’ matching. In this study, when the matching scheme presented ‘one-to-many’ mapping, the frequency of the USAS tag was equally distributed among all of the possible matching domains/fields. So, for example, in Step 1, conceptual domain SUBSTANCES AND MATERIALS: SOLID (78%) was equally distributed between FOOD (39%), and OTHER (39%). Though this clearly leads to an approximation, it seemed the only possible solution,

since manual tags refer to the relationship that exists between the key word (*chocolate* or *wine*) and the rest of the sentence, while automatic tags describe individual words, regardless of the key word. Manually looking at individual concordances in order to recreate the relationship to the key word would have been off the point in this case, as the aim of the study is precisely to investigate and assess automated procedures.

Research design

For the sake of clarity, this section summarises the different preparatory and analytical steps performed in this study. Collecting the elicited data and tagging them (first manually, and then automatically with Wmatrix), as well as extracting 10,000 sentences around each node word from UKWAC and tagging them with Wmatrix are all considered preparatory phases and are summarised in Table 6.

Wmatrix provided four different types of lists, two tagged (semantic word list and semantic keyword list) and two untagged (raw word list and raw keyword list). Manual coding, on the other hand, provided two sets of lists, respectively showing the frequency of semantic fields and of conceptual domains.

Table 6. Summary of preparatory phases.

	<i>Description</i>	<i>Format</i>	<i>Method / Tool</i>
<i>Preparatory phases</i>	Collecting elicited data	questionnaires	
	Coding elicited data manually (at sentence level)	Excel file	manually
	Tagging elicited data automatically (at word level)	Text file	Wmatrix
	Extracting concordance lines, and full sentences from UKWAC	Text file	The Sketch Engine
	Tagging sentences from UKWAC	Text file	Wmatrix

Using the output lists from Wmatrix, the analyses in Table 7 were performed. First of all, automatic tagging was compared to manual tagging, at the two levels of conceptual domain and semantic field. Finally, taking advantage of automatic tagging, the elicited data were compared to the Web data.

Table 7. Summary of analytical phases.

	<i>Step</i>	<i>Description</i>	<i>Method / Tool</i>
<i>Analyses</i>	STEP 1	Elicited data: manual vs. automatic tagging (conceptual domains)	Spearman Rank Correlation Coefficient
	STEP 2	Elicited data: manual vs. automatic tagging (semantic fields)	Spearman Rank Correlation Coefficient
	STEP 3	Elicited data vs. Web data: automatic tagging (semantic fields)	Spearman Rank Correlation Coefficient

These steps were carried out separately for each of the two node words, *chocolate* and *wine*. The analyses and their results are detailed and discussed in the following section.

Results

A first goal of this study was to compare the potential of manual coding to automatic tagging. This comparison was performed at the level of conceptual domains (Step 1), as well as of semantic fields (Step 2). Manual coding will be here considered as a sort of control situation. Another goal of the study was to compare elicited data to Web data (Step 3). At this level of analysis, comparison was performed using automatic tagging only, therefore no conversion was necessary. The results of these comparisons are described in the following paragraphs.

Manual vs. automatic tagging – Step 1

Comparison between manual coding and automatic tagging was first performed at the level of conceptual domains (superordinate, broader categories). To this end, the matching scheme described in the *Matching manual tags to automatic tags* Section above was applied to the top 30 items in the semantic frequency list and in the semantic keyword list of the elicited data as offered by Wmatrix, excluding items in Table 5. Thirty is an arbitrary number selected out of convenience, as a consequence of the fact that prominent semantic fields and domains were expected to emerge through frequency. As already mentioned, when the matching scheme presented ‘one-to-many’ mapping, the frequency of the USAS tag was equally distributed among all of the possible matching conceptual domains (e.g. SUBSTANCES AND MATERIALS: SOLID (78%) = FOOD (39%), and OTHER (39%)).

As an intermediate step between manual tagging (sentence-based) and semantic tagging (word-based), it was decided to consider also the top 30 items of the raw frequency list and of the keyword list, as this allowed us to apply manual tagging on the basis of individual words. Therefore, the top 30 semantic items in the lists (excluding the node word) were manually mapped to one or more of the conceptual domains described in the Codebook. Thus, for example, in the Chocolate set of data, the word *white* was matched to FEATURE (as it could either indicate a colour or a type of chocolate), and the word *Cadbury* was matched to both FOOD and GEOGRAPHY (as it refers to a manufacturing industry, but also to a well recognizable geographical origin: England).

Tables 8 and 9 summarise the conceptual domains emerging in the elicited data with the five different tagging procedures, for the node words *chocolate* and *wine* respectively. The numbers in the first column indicate ranking. Numbers in parenthesis are percentages, rounded to the second decimal.

Table 8. Conceptual domains in the Chocolate data.

	Manual tagging	Raw frequency list	Semantic frequency list	Raw keyword list	Semantic keyword list
1	food (29.85)	feelings & emotions (2.64)	food (19.96)	feelings & emotions (3.08)	food (20.50)
2	features (25.61)	features (2.13)	features (4.65)	food (2.72)	life (5.86)
3	feelings & emotions (23.38)	food (1.85)	feelings & emotions (3.68)	features (2.37)	features (4.58)
4	health & body (12.30)	health & body (1.15)	events (1.97)	health & body (0.66)	feelings & emotions (3.38)
5	events (11.29)	events (0.77)	people (1.76)	events (0.41)	events (1.97)
6	people (9.44)	people (0.54)	health & body (1.22)	geography (0.12)	people (0.82)

7	geography (4.67)	geography (0.25)	geography (0.95)		health & body (0.53)
8	culture (2.55)				
9	environment (2.39)				
10	ceremonies (1.96)				
11	loss & damage (1.1)				
12	imagination (0.8)				
13	life (0.48)				
14	comparison (0.37)				
15	sports (0.05)				

Table 9. Conceptual domains in the Wine data.

	Manual tagging	Raw frequency list	Semantic frequency list	Raw keyword list	Semantic keyword list
1	features (31.92)	feature (5.53)	feature (5.36)	feature (5.76)	feature (4.26)
2	food (24.39)	food (4.20)	food (3.87)	food (3.19)	food (2.90)
3	people (13.87)	people (0.71)	events (3.28)	events (0.41)	events (2.61)
4	events (13.31)	health & body (0.65)	health & body (1.93)	comparison (0.40)	health & body (2.22)
5	feelings & emotions (10.68)	events (0.63)	people (1.89)	geography (0.31)	geography (1.30)
6	health & body (9.39)	feelings & emotions (0.50)	geography (1.56)	people (0.23)	people (0.92)
7	geography (7.79)	comparison (0.21)	feelings & emotions (1.33)	feelings & emotions (0.14)	feelings & emotions (0.84)
8	comparison (2.89)	geography (0.16)	comparison (0.58)	health & body (0.12)	comparison (0.29)
9	ceremonies (2.53)		ceremonies (0.22)		
10	environment (1.29)				
11	culture (0.93)				
12	loss & damage (0.52)				
13	life (0.52)				
14	imagination (0.05)				

As the tables show, the same conceptual domains were found in the questionnaires and in the top 30 semantic words of the different word lists and keyword lists, but with some disparities in the rankings. In order to decide whether the observed similarities and differences were significant, and which of the four methodological approaches better describes the given population, Spearman's Rank Correlation Coefficient was applied. This is a non-parametric (i.e. distribution-free) test, appropriate to ordinal scales, which uses ranks of the x and y variables, rather than data (Fowler, Cohen, & Jarvis, 1998: 138-141). Spearman's r "describes the overlap of the variance of ranks" (Arndt, Turvey, & Andreasen, 1999, p. 104). Correlation was performed using SPSS.

A positive type of correlation was found in each of the four cases. However, different methodological approaches gave slightly different results when applied to the chocolate or the wine data. For *chocolate*, the strongest correlation was found using the semantic frequency list ($r = 0.929$, $P < 0.01$), immediately followed by the raw frequency list ($r = 0.905$, $P < 0.01$) and the raw keyword list ($r = 0.886$, $P < 0.01$). The semantic keyword list, on the other hand, showed weak correlation ($r = 0.429$, $P < 0.05$). For *wine*, the strongest correlation was found using the raw

frequency list ($r = 0.933$, $P < 0.01$), immediately followed by the semantic frequency list ($r = 0.883$, $P < 0.01$), the semantic keyword list ($r = 0.817$, $P < 0.01$), and finally the raw keyword list ($r = 0.683$, $P < 0.05$). On the basis of these results, it could tentatively be suggested that the most representative methods seem to be frequency list and semantic frequency list, as they showed strong correlation in both datasets. Interestingly, the raw keyword list never did show strong correlation results, while the semantic keyword list performed very differently in the two experiments, with high correlation in one case and very low correlation in the other. However, further investigation with a wider number of datasets is necessary, before statistically sound conclusions can be drawn.

Manual vs. automatic tagging – Step 2

At the level of semantic fields, comparison between manual and automatic tagging was performed using (a) the semantic keyword lists, and (b) the semantic word lists. In experiment (a), the conversion scheme described in Section 3 was applied to the positive keywords; next, the lists of semantic fields obtained applying the conversion scheme were matched to the semantic fields lists of the manual tagging (Tables A and B in the Appendix), and SPSS was used to perform the correlation by applying Spearman's Rank Correlation Coefficient. A significant positive correlation was found in both cases, with the result for *chocolate* falling in the strong range ($r = 0.703$ at $P < 0.01$) and that for *wine* in the modest range ($r = 0.486$ at $P < 0.01$).

A similar procedure was used with the semantic word lists, considering only the top 50 items in the list (Tables C and D in the Appendix). Both the chocolate and the wine datasets showed positive correlation in the medium range, with results for *chocolate* being $r = 0.505$ at $P < 0.01$, and for *wine* $r = 0.558$ at $P < 0.01$. Interestingly, while use of the semantic keyword list led to results that are dependent on the dataset, use of the semantic word list provided very similar results for the two datasets. This seems to confirm that the semantic frequency list is more representative than the semantic keyword list. Comparison using the whole semantic word list was not performed in the current study, however – given the results obtained in Step 3 and described in the following paragraphs – correlation could be expected to be even higher.

Elicited data vs. Web data - Step 3

Finally, automatic tagging of elicited data was contrasted to automatic tagging of Web data. Comparison was fairly straightforward, as no tagging conversion scheme was required. The semantic word lists of the elicited data were aligned to the semantic word lists of the Web data (Tables E and F in the Appendix); correlation was assessed using SPSS and by applying Spearman's Rank Correlation Coefficient.

For the sake of experimentation, correlation was computed in three different ways: (1) using the whole semantic frequency lists, (2) using the top 100 items in the lists; and (3) using the top 50 items. All the six cases (three for chocolate and three for wine) showed interesting positive correlation between the elicited and the Web data, the strength of the correlation decreasing as the number of items considered decreased. In fact, for both *wine* and *chocolate*, comparison of whole lists showed strong correlation (*chocolate*: $r = 0.790$, $P < 0.01$; *wine*: $r = 0.791$, $P < 0.01$), comparison using the top 100 items showed medium correlation (*chocolate*: $r = 0.492$, $P < 0.01$; *wine*: $r = 0.437$, $P < 0.01$), while comparison using the top 50 items showed low-medium correlation (*chocolate*: $r = 0.341$, $P < 0.01$; *wine*: $r = 0.548$, $P < 0.01$).

Conclusions

The analyses performed in this study aimed at assessing different methodological approaches to automatic semantic tagging in the analysis of cultural traits, in order to lay some foundation blocks in the definition of a method for the use of medium-large corpora for the extraction of EMUs.

Because of their dimension, large corpora can only be tagged in their entirety by automatic tagging systems. A medium-sized corpus like the ones used in this study (about 10,000 sentences and 400,000 words each), could, if necessary, be tagged manually; however, such a tagging process would take several weeks, if we consider that completing the coding process of each of the elicited datasets in this study (only about 1,900 sentences and 13,000 words) required little more than a week and the work of two people. For this reason it seemed important to establish whether automatic tagging could provide the same results as manual tagging. This was done in Steps 1 and 2, by comparing the results of automatic and semantic tagging on the elicited data. Once it was established that the semantic word list obtained with automatic tagging was sufficiently representative of the range of EMUs that emerged from manual tagging, it was possible to proceed to a comparison between the elicited and the Web data (Step 3).

Furthermore, a quick Google search for the word “semantic tagger” shows that for the time being, few languages can benefit from an automatic semantic tagger, and the existing taggers seem to be based on different semantic tagsets. This means that, if one wanted to perform cross-cultural comparisons, either manual tagging and aligning of manual tags to automatic ones, or aligning of two different tagsets will still have to be performed. Therefore, in the prospect of a future need to contrast British EMUs to Italian EMUs, this paper investigated also correspondence between the full corpus and different types of subsets of the corpus (top 30 words in word list; top 30, top 50, and top 100 items in semantic frequency list; top 30 items in frequency and semantic keyword lists), in the hope of finding shortcuts to the most frequent EMUs, without tagging the whole corpus.

In this respect, Step 1 – i.e. comparison between manual coding of the whole corpus and the top 30 items in the raw and semantic word lists and keyword lists, performed at the level of conceptual domains – was particularly important. Its results, in fact, suggest that the top 30 items in the raw frequency list and in the semantic frequency list could both be considered representative of the most frequent EMUs, as they showed strong correlation to the results of manual tagging, in both the *chocolate* and the *wine* datasets. On the other hand, the top 30 items in the raw keyword list never showed strong correlation, while the top 30 items in the semantic keyword list performed differently in the two datasets, with high correlation in one case and very low correlation in the other. Confirmation that the semantic word list is more adequate than the semantic keyword list for comparing results of different taggings was also found in Step 2, where comparison between manual coding and automatic tagging was performed at the level of semantic fields. In fact, at this stage of the investigation, when the semantic word list was used, the same type of positive correlation was found between manual and automatic tagging in both the *chocolate* and the *wine* datasets, while use of the semantic keyword list led to correlation results that were rather different in the two datasets.

The fact that the EMUs that emerged as most prominent through manual coding also emerged in the top 30 (or in some cases 50) items in the lists is in keeping with both Szalay and Maday's theory (1973) which measures the dominance component of EMUs in terms of

frequency, and with Fleischer's (1998) which considers frequency as an indication of conventionalisation. Furthermore, it is interesting to notice that the raw and semantic keyword lists – obtained by contrasting the elicited data word list to that of the BNC – provided data subsets that were no longer representative of the whole datasets.

Finally, comparison between the elicited data and the Web data (Step 3) showed an interesting positive correlation which fell in the middle range when the whole semantic frequency list was used. This result, though not excellent, is highly encouraging, given the limited amount of information that is available about and the control that it is possible to have over the contents of a general Web corpus created with spidering tools. In fact, for the moment, nothing can guarantee that the UKWAC corpus includes only texts written by British native speakers. Furthermore, despite all the efforts by the compilers of the UKWAC corpus, a small amount of noise was still detectable in the sentences extracted from it. Also, it is highly possible that the elicited data and the Web data used in this study mirrored rather different populations; indeed the selection of participants in the survey provided a population that is quite unbalanced at least in terms of age and level of education, as it is mostly composed of university students (70%). Finally, going through the sentences extracted from the UKWAC, it can easily be seen that a noticeable percentage of them are actually extracts from recipes. These belong to a different level of culture, the level Hall (1989) calls 'technical', while EMUs are more likely to emerge at the informal level.

Though further investigation with a wider number of datasets is necessary, before a final procedure for the automatic extraction of EMUs from large Web corpora can be designed, I hope that this paper may offer an interesting starting point for future reflection. In fact, the results of this study are highly encouraging as regards the use of a semantic tagger like Wmatrix for the automatic extraction of EMUs and also as regards the possibility of using in cultural analysis general Web corpora created using spidering tools.

References

- Archer, D., Wilson, A., & Rayson, P. (2002). Introduction to the USAS category system. Benedict project report, October 2002.
Retrieved 10 May 2009 from <http://ucrel.lancs.ac.uk/usas/usas20guide.pdf>
- Archer, D., Rayson, P., Piao, S., & McEnery, T. (2004). Comparing the UCREL semantic annotation scheme with lexicographical taxonomies. *Proceedings of the EURALEX-2004 Conference* (pp. 817-827). Lorient, France.
- Arndt, S., Turvey, C., & Andreasen, N.C. (1999). Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation. *Journal of Psychiatric Research*, 33, 97-104.
- Baroni, M., & Kilgarrieff, A. (2006). Large linguistically processed web corpora for multiple languages. *Proceedings of EACL*. Trento.
- Baroni, M., & Bernardini, S. (2004). Bootcat: Bootstrapping corpora and terms from the web. *Proceedings of the Fourth Language Resources and Evaluation Conference* (pp. 1313-1316). Lisbon, Portugal.
- Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2008). The WaCky Wide Web: A collection of very large, linguistically processed Web-crawled corpora. *Journal of Language Resources and Evaluation*, 43 (3), 209-226.

- Baroni, M., & Ueyama, M. (2006). Building general- and special-purpose corpora by Web crawling. *Proceedings of 13th NIJL International Symposium, Language Corpora: Their Compilation and Application* (pp. 31-40). Tokyo, Japan.
- Bianchi, F. (2007). The cultural profile of chocolate in current Italian society: A corpus-based pilot study. *ETC – Empirical Text and Culture Research*, 3, 106-120.
- Bianchi, F. (2009). From language to culture: Looking for quantitative parameters for assessing and comparing cultures through corpora. In M. Dossena, D. Torretta, & A.M. Sportelli (Eds), *Forms of migration – Migration of forms: Proceedings of the 23rd AIA Conference: Bari (Italy), 20-22 September 2007* (pp. 227-241). Bari: Progreedit.
- Cavalli-Sforza, L.L. (1996). *Geni, popoli e lingue* [Genes, peoples, and languages]. Milano: Adelphi.
- Crystal, D. (2003). *English as a global language*. Cambridge: Cambridge University Press.
- Evert, S. (2008). Corpora and collocations. In A. Lüdeling, & M. Kytö (Eds), *Corpus linguistics. An international handbook. Vol. 2* (pp. 1212-1248). Berlin, New York: Mouton de Gruyter.
- Fairclough, N. (2003). *Analysing discourse*. London & New York: Routledge.
- Fleischer, M. (1998). Concept of the ‘Second Reality’ from the perspective of an empirical systems theory on the basis of radical constructivism. In G. Altman, & W.A. Koch (Eds), *Systems. New paradigms for the human sciences* (pp. 423-460). Berlin-NewYork: De Gruyter.
- Fleischer, M. (2002). Das Image von Getränken in der polnischen, deutschen und französischen Kultur [The image of drinks in the Polish, German and French cultures]. *ETC: Empirical Text and Culture Research*, 2, 8-47.
- Fletcher, W. (2004). Making the Web more useful as a source for linguistic corpora. In U. Cornor, & T. Upton (Eds), *Corpus linguistics in North America 2002* (pp. 191-205). Amsterdam: Rodopi.
- Fowler, J., Cohen L., & Jarvis P. (1998). *Practical statistics for field biology (2nd ed.)*. Chichester: Wiley.
- Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: CLAWS4. In R. Garside, G. Leech, & A. McEnery (Eds), *Corpus annotation: Linguistic information from computer text corpora* (pp. 102-121). London: Longman.
- Greenberg, J.H. (1960). Concerning inferences from linguistic to nonlinguistic data. In H. Hoiyer (Ed.), *Language in culture* (pp. 3-20). Chicago: University of Chicago.
- Hair, J.F., Bush, R.P., & Ortinau, D.J. (2009). *Marketing research in a digital information environment. Fourth edition*. Boston: McGraw-Hill.
- Hall, E.T. (1989). *Beyond culture*. New York: Doubleday.
- Halliday, M.A.K. (1978). *Language as social semiotics. The social interpretation of language and meaning*. Open University. London: Arnold Ed.
- Kilgarriff, A., Rychly, P., Smrz, P., & Tugwell, D. (2004). The Sketch Engine. *Euralex XI Proceedings*, Lorient, France, 105–11.
- Leech, J. & Fallon, R. (1992). Computer corpora: What do they tell us about culture?. *ICAME Journal*, 16, 29-50.
- Lotman, J.M. (1993). *La cultura e l'esplosione: prevedibilità e imprevedibilità*. Milano: Feltrinelli.
- Malinowski, B. (1923). The problem of meaning in primitive languages. In C.K. Ogden, & I.A. Richards (Eds), *The meaning of meaning: A study of the influence of language upon thought and of science of symbolism* (pp. 296-336). New York: Harcourt Brace and Co.
- McArthur, T. (1981). *Longman lexicon of contemporary English*. Longman, London.

- Mitchell, T.L., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., & Just, M.A. (2008). Predicting human brain activity associated with meanings of nouns. *Science*, 230, 1191-1195.
- Muntz, R. (2001). Evidence of Australian cultural identity through the analysis of Australian and British corpora. In P. Rayson, A. Wilson, T. McEnery, A. Hardie, & S. Khoja (Eds) *Proceedings of the Corpus Linguistics 2001 Conference. UCREL Technical Papers 13 (Special Issue)* (pp. 393-399). Lancaster University.
- Murphy B., Baroni, M., & Poesio, M. (2009). EEG responds to conceptual stimuli and corpus semantics. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)* (pp. 619-627). East Stroudsburg PA: ACL.
- Neuedorf, K.A. (2002). *The content analysis guidebook*. Thousand Oaks - London - New Delhi: Sage Publications.
- Nobis, A. (1998). Self-organisation of culture. In G. Altmann, & W.A. Koch (Eds), *Systems. New paradigms for the human sciences* (pp. 461-478). Berlin-NewYork: De Gruyter.
- Oakes, M. P. (2003). Contrasts between US and British English of the 1990s. In E. H. Oleksy, & B. Lewandowska-Tomaszczyk (Eds), *Research and scholarship in integration processes* (pp. 213-22). Lodz: University of Lodz Press.
- Partington, A. (2004). Utterly content in each other's company. Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9, 131-156.
- Piao, S.L., Rayson, P., Archer, D., & McEnery, T. (2004). Evaluating lexical resources for a semantic tagger. *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC 2004), May 2004, Lisbon, Portugal, Volume II* (pp. 499-502).
- Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13 (4), 519-549.
- Rayson, P., Archer, D., Piao, S., & McEnery, A.M. (2004). The UCREL semantic analysis system. *Proceedings of the Beyond Named Entity Recognition Semantic Labeling for NLP Tasks Workshop, Lisbon, Portugal* (pp. 7-12).
- Sapir, E. (edited by D.G. Mandelbaum) (1949). *Selected writings in language, culture and personality*. Berkeley: University of California Press.
- Schmid, H.J. (2003). Do women and men really live in different cultures? Evidence from the BNC. In A. Wilson, P. Rayson and T. McEnery (eds) *Corpora by the Lune. A festschrift for Geoffrey Leech* (pp. 185-221). Frankfurt: Peter Lang.
- Sharoff, S. (2006). Creating general-purpose corpora using automated search engine queries. In M. Baroni, & S. Bernardini (Eds), *WaCky! Working papers on the Web as Corpus. Gedit*, Bologna: Gedit.
- Szalay, L.B., & Maday, B.C. (1973). Verbal associations in the analysis of subjective culture. *Current Anthropology*, 14(1-2), 33-42.
- Tannen, D. (1990). *You just don't understand*. New York: Ballantine Books.
- Ueyama, M. (2006). Creation of general-purpose Japanese Web corpora with different search engine query strategies. In M. Baroni, & S. Bernardini (Eds), *WaCky! Working papers on the Web as Corpus*. Bologna: Gedit.
- Whorf, B.L. (1956). *Language, thought, and reality*. New York: John Wiley & Sons, and The Technology Press of M.I.T.
- Wilson, A., & Rayson, P. (1993). Automatic content analysis of spoken discourse: a report on work in progress. In C. Souter, & E. Atwell (Eds) *Corpus based computational linguistics* (pp. 215-226). Amsterdam: Rodopi.

Appendix

Table A
Chocolate: comparison of semantic fields using semantic keyword list (Step 2a)

Semantic field	manual tagging	automatic tagging	Semantic field	manual tagging	automatic tagging	Semantic field	manual tagging	automatic tagging
animals	17	38	friendship	8	20	physical propert.	14	0
artistic production	46	0	Future	2	0	pleasure	26	40
bakery/cooking	54	0	Gay	2	0	posh	1	0
beauty	19	3	genuine	0	0	price	27	20
body	95	138	geo locations	80	66	product/shape	184	29
bribing	4	0	Gift	35	67	quality/type	177	61
ceremonies	1	0	guilt	9	5	quantity	56	78
children	42	0	happiness	116	99	recipe	46	0
colour	26	132	health	84	97	relax	14	5
comfort	20	5	hiding	2	0	religion	9	0
comparison	7	11	history	2	0	royalty	3	0
competitiveness	0	0	house	4	0	seduction	9	0
composition	68	54	language	3	0	senses	14	0
culture	2	0	law	4	0	sex	23	0
desire	106	103	loneliness	1	0	sharing/society	18	0
dirt	18	0	love	14	0	sports	1	0
dream	4	6	manufacturing	49	29	spreading	8	0
drink	47	223	medicine	34	69	studying/intellect	0	0
drugs & addiction	13	10	memory	8	0	surprise	1	0
economy	10	2	men	27	0	sweet	22	0
energy	21	11	mood	19	5	taste/smell	125	139
event	79	112	nature	2	0	tech	4	0
existence	7	5	packaging	15	0	theft	5	12
fair trade	10	0	party	1	0	time	35	11
family	10	0	passion	48	23	transaction	54	85
fantasy/magic	11	0	peace	8	5	war	2	0
food	115	1845	people	16	82	women	51	23
freedom	1	0	persuasion	0	0	work	5	0

Table B
Wine: comparison of semantic fields using semantic keyword list (Step 2a)

Semantic field	manual tagging	automatic tagging	Semantic field	manual tagging	automatic tagging	Semantic field	manual tagging	automatic tagging
age	20	33	freedom	2	2	physical properties	41	40
animals	2	0	friendship	30	40	pleasure	6	0
artistic production	9	13	future	1	0	posh	38	0
bakery/cooking	35	0	gay	2	0	price	101	87
beauty	1	114	genuine	2	0	product/shape	35	0
body	10	0	geo locations	141	153	quality/type	228	153
bribing	1	0	gift	21	47	quantity	64	75
ceremonies	11	0	guilt	1	0	recipe	61	0
children	5	0	happiness	57	55	relax	36	0
colour	23	156	health	129	13	religion	30	86
comfort	6	0	hiding	3	0	royalty	0	0
comparison	56	70	history	3	0	seduction	2	0
competitiveness	0	0	holidays	7	0	senses	4	1
composition	66	58	house	3	14	serving	15	0
confidence	2	4	illicit drugs	0	2	sex	1	12
culture	4	0	language	14	0	sharing/society	39	9
desire	48	0	law	1	0	sports	0	0
dirt	17	0	loneliness	0	0	spreading	10	0
dream	0	5	love	11	0	storage	32	21
drink	93	1888	manufacturing	36	0	studying/intellect	5	2
driving	5	0	medicine	42	130	surprise	1	0
drugs & addiction	5	0	memory	5	28	sweet	7	0
economy	3	0	men	38	0	taste/smell	117	115
energy	0	0	mood	5	8	tech	0	0
event	23	0	nature	3	19	theft	2	14
excessive drinking	82	81	packaging	36	0	time	44	108
existence	9	3	party	17	0	transaction	30	62
fair trade	0	0	passion	16	7	war	1	0
family	35	0	peace	3	28	women	52	19
fantasy/magic	1	0	people	10	167	work	15	3
food	100	282	persuasion	0	0			

Table C
Chocolate: comparison of semantic fields using semantic word list (Step 2b)

Semantic field	manual tagging	automatic tagging		Semantic field	manual tagging	automatic tagging
product/shape	184	10		physical propert.	14	0
quality/type	177	86.5		relax	14	0
taste/smell	125	181.1		senses	14	0
happiness	116	90		drugs and addiction	13	0
food	115	1845		fantasy/magic	11	0
desire	106	45.5		economy	10	0
body	95	0		fair trade	10	0
health	84	93.6		family	10	35
geo locations	80	66		guilt	9	0
event	79	0		religion	9	112
composition	68	17		seduction	9	0
quantity	56	165		friendship	8	0
bakery/cooking	54	0		memory	8	0
transaction	54	85		peace	8	0
women	51	0		spreading	8	0
manufacturing	49	0		comparison	7	28
passion	48	217		existence	7	0
drink	47	187		theft	5	0
artistic production	46	0		work	5	0
recipe	46	0		bribing	4	0
children	42	0		dream	4	0
gift	35	27.5		house	4	0
time	35	41		law	4	0
medicine	34	95.5		tech	4	0
men	27	0		language	3	13
price	27	0		royalty	3	0
colour	26	41		culture	2	0
pleasure	26	0		future	2	0
sex	23	0		gay	2	0
sweet	22	0		hiding	2	0
energy	21	0		history	2	0
comfort	20	0		nature	2	0
beauty	19	70		war	2	0
mood	19	0		ceremonies	1	0
dirt	18	0		freedom	1	0
sharing/society	18	0		loneliness	1	0
animals	17	38		party	1	0
people	16	147		posh	1	0
packaging	15	10		sports	1	0
love	14	0		surprise	1	0

Table D
. Wine: comparison of semantic fields using semantic word list (Step 2b)

semantic field	manual tagging	automatic tagging		semantic field	manual tagging	automatic tagging
age	20	0		language	14	0
animals	2	0		law	1	0
artistic production	9	0		love	11	0
bakery/cooking	35	0		manufacturing	36	0
beauty	1	82		medicine	42	99
body	10	0		memory	5	28
bribing	1	0		men	38	0
ceremonies	11	0		mood	5	0
children	5	0		nature	3	9.3
colour	23	306		packaging	36	26.6
comfort	6	0		party	17	0
comparison	56	64		passion	16	197
composition	66	22		peace	3	14
confidence	2	0		people	10	167
culture	4	0		physical properties	41	0
desire	48	24.5		pleasure	6	0
dirt	17	0		posh	38	0
drink	93	1869		price	101	41
driving	5	0		product/shape	35	0
drugs and addiction	5	0		quality/type	228	12.5
economy	3	0		quantity	64	155.5
event	23	0		recipe	61	0
excessive drinking	82	79		relax	36	14
existence	9	0		religion	30	86
family	35	52		seduction	2	0
fantasy/magic	1	0		senses	4	0
food	100	282		serving	15	26.6
freedom	2	0		sex	1	0
friendship	30	20		sharing/society	39	0
future	1	0		spreading	10	0
gay	2	0		storage	32	0
genuine	2	0		studying/intellect	5	0
geo locations	141	152.3		surprise	1	0
gift	21	23.5		sweet	7	0
guilt	1	0		taste/smell	117	156.2
happiness	57	55		theft	2	0
health	129	102.2		time	44	129
hiding	3	0		transaction	30	62
history	3	0		war	1	0
holidays	7	0		women	52	0
house	3	0				

Table E
 . Step 3: Chocolate: Correlation table using semantic word list (Step 3)

USAS tag	Web	elicited		USAS tag	Web	elicited		USAS tag	Web	elicited
A1.1.1	3674	238		H4	520	11		S1.2.1-	14	0
A1.1.1-	4	0		H4-	6	0		S1.2.1+	75	2
A1.1.2	222	10		H5	540	10		S1.2.2-	17	0
A1.2	36	0		I1	941	3		S1.2.2+	20	0
A1.2-	4	0		I1.1-	34	1		S1.2.3-	2	1
A1.2+	90	1		I1.1---	3	0		S1.2.3+	39	0
A1.3-	4	0		I1.1	236	1		S1.2.3+++	1	0
A1.3+	57	0		I1.1+	202	5		S1.2.4-	16	0
A1.4-	10	2		I1.1++	2	1		S1.2.4	1	0
A1.4	110	1		I1.1+++	31	0		S1.2.4+	149	1
A1.4+	42	0		I1.2	202	1		S1.2.5---	1	0
A1.5	1	0		I1.3---	6	0		S1.2.5-	26	2
A1.5.1	552	26		I1.3-	104	11		S1.2.5+	87	1
A1.5.1-	4	0		I1.3--	10	1		S1.2.5++	4	0
A1.5.1+	3	0		I1.3	316	1		S1.2.5+++	2	0
A1.5.2-	12	0		I1.3+	28	9		S1.2.6---	1	0
A1.5.2+	51	0		I2	3	0		S1.2.6-	52	5
A1.6	14	0		I2.1	543	1		S1.2.6+	20	0
A1.7-	137	7		I2.2	2009	85		S2	981	82
A1.7+	144	0		I3.1	441	2		S2.1	301	23
A1.8-	47	0		I3.1-	15	0		S2.2	360	10
A1.8+	876	20		I3.2	29	0		S3.1	254	20
A1.9	74	0		I3.2-	2	0		S3.1-	7	0
A1.9-	3	0		I3.2+	29	0		S3.2	297	11
A10-	342	8		I4	397	3		S3.2-	1	0
A10---	1	0		K1	581	4		S3.2+	3	0
A10+	765	19		K2	585	3		S4	764	35
A11.1-	48	1		K3	135	1		S4-	1	0
A11.1--	1	0		K4	218	0		S4T1.1.1	1	0
A11.1+	452	8		K5	19	0		S5-	115	0
A11.1++	120	8		K5.1	460	2		S5	2	0
A11.1+++	23	8		K5.2	133	0		S5+	910	7
A11.2-	12	8		K6	125	0		S5+++	5	0
A11.2+	70	8		L1-	168	4		S6-	221	3
A12-	177	5		L1	240	5		S6+	663	25
A12+	130	3		L1+	129	5		S7.1-	72	4
A12++	14	0		L2	1933	38		S7.1--	8	0
A12+++	1	0		L2-	2	0		S7.1	13	0
A13	140	6		L3	2149	3		S7.1+	873	5
A13.1	221	3		M1	2087	64		S7.1++	1	0
A13.2	385	9		M2	1525	22		S7.2-	3	0
A13.3	1149	67		M3	593	9		S7.2+	51	1
A13.4	289	3		M4	322	2		S7.3	17	1
A13.5	129	5		M5	121	0		S7.3+	34	0
A13.6	219	5		M6	2977	38		S7.4-	56	3
A13.7	88	5		M7	1093	6		S7.4	2	0
A14	702	18		M8	201	3		S7.4+	215	10

USAS tag	Web	elicited		USAS tag	Web	elicited		USAS tag	Web	elicited
A15-	105	1		N1	1327	20		S8-	121	6
A15	1	0		N2	74	0		S8+	1355	25
A15+	37	0		N3	107	9		S9	975	112
A15++	1	0		N3.1	29	0		S9-	1	0
A2.1-	38	2		N3.2-	465	6		T1	337	10
A2.1	4	0		N3.2--	19	0		T1.1	92	1
A2.1+	844	37		N3.2---	4	0		T1.1.1	411	17
A2.2	915	24		N3.2	109	5		T1.1.2	786	29
A2.2-	2	0		N3.2+	488	68		T1.1.3	972	9
A3	15	0		N3.2++	29	0		T1.2	463	5
A3-	13	0		N3.2+++	129	3		T1.3	2816	41
A3+	5309	586		N3.3-	63	0		T1.3-	42	1
A4.1	910	41		N3.3--	11	0		T1.3+	99	3
A4.2-	32	0		N3.3---	8	0		T1.3++	23	0
A4.2--	1	0		N3.3	461	2		T1.3+++	5	0
A4.2+	441	7		N3.3+	42	1		T2-	532	17
A4.2++	1	0		N3.3++	3	0		T2+	518	8
A5.1---	20	1		N3.4	206	0		T2++	351	5
A5.1-	156	3		N3.4-	1	0		T2+++	45	2
A5.1--	16	0		N3.4+	3	2		T3---	111	1
A5.1	193	0		N3.5--	2	1		T3-	1021	7
A5.1-	0	30		N3.5	489	4		T3--	147	0
A5.1+	1234	102		N3.5-	8	0		T3	120	3
A5.1++	140	0		N3.5+	32	4		T3+	233	4
A5.1++	0	26		N3.6-	11	0		T3++	51	1
A5.1+++	572	44		N3.6	55	0		T3+++	17	0
A5.1++++++	2	0		N3.6+	3	0		T4--	51	0
A5.2-	93	0		N3.7-	103	2		T4-	52	0
A5.2+	232	4		N3.7---	2	1		T4	4	0
A5.3-	51	3		N3.7	186	1		T4+	50	0
A5.3+	81	0		N3.7--	3	0		T4++	1	0
A5.4-	73	2		N3.7+	368	6		W1	302	19
A5.4+	147	8		N3.7++	32	0		W2-	376	50
A5.4+++	3	0		N3.7+++	17	0		W2--	11	0
A6	4	0		N3.8-	67	0		W2---	4	0
A6.1-	883	28		N3.8	39	1		W2	183	1
A6.1	55	0		N3.8+	178	6		W3	617	6
A6.1+	287	8		N3.8++	8	0		W4	192	1
A6.1++	1	0		N3.8+++	13	0		W5	40	2
A6.1+++	132	1		N4	1163	11		X1	19	0
A6.2-	170	3		N4-	1	0		X2	39	6
A6.2+	467	15		N5---	165	0		X2.1	706	98
A6.2+++	4	0		N5	1466	63		X2.1-	1	0
A6.3+	355	4		N5-	436	8		X2.2-	51	0
A6.3++	2	0		N5--	42	1		X2.2	12	0
A7	251	1		N5.1-	385	3		X2.2+	713	17
A7-	37	0		N5.1	5	0		X2.2++	2	0
A7+	1998	189		N5.1+	1647	39		X2.2+++	2	0

USAS tag	Web	elicited		USAS tag	Web	elicited		USAS tag	Web	elicited
A7+++	31	0		N5.1++	2	2		X2.3+	97	1
A8	275	4		N5.1+++	7	0		X2.4	381	2
A8+	2	0		N5.2+	218	57		X2.5-	35	0
A9-	1443	57		N5+	1324	54		X2.5+	81	1
A9	26	1		N5++	1236	18		X2.6-	74	1
A9+	2584	105		N5+++	110	10		X2.6	5	0
B1	1646	69		N5++++	1	0		X2.6+	133	2
B2-	561	53		N6-	232	7		X3	37	0
B2	78	3		N6---	1	0		X3.1	1007	76
B2+	129	3		N6	86	11		X3.1+	248	38
B2++	12	0		N6+	364	24		X3.2	278	5
B2+++	2	1		N6+++	107	6		X3.2-	32	0
B3	398	9		N6++++	1	0		X3.2+	13	0
B4	415	3		O1	912	34		X3.3	61	2
B4-	1	0		O1.1	1760	12		X3.4	466	12
B5	1007	6		O1.2	800	1		X3.4-	6	0
B5-	14	0		O1.2-	95	0		X3.4+	9	0
C1	797	7		O1.3	104	0		X3.5	142	7
E1	89	20		O1.3-	2	0		X3.5+	1	0
E1-	2	0		O2	4782	45		X4.1	383	3
E2-	38	12		O3	167	0		X4.2	388	4
E2	3	0		O4.1	1066	12		X5.1-	55	0
E2+	760	187		O4.2-	139	10		X5.1+	76	0
E2++	0	11		O4.2	11	0		X5.1++	2	0
E2++	59	0		O4.2+	1060	70		X5.2-	55	1
E2+++	167	30		O4.2++	28	2		X5.2+	461	22
E3-	260	5		O4.2+++	5	3		X5.2++	5	0
E3+	167	15		O4.2++++	2	0		X5.2+++	4	0
E4.1-	133	6		O4.3	2667	82		X6	17	0
E4.1+	525	90		O4.4	566	16		X6+	82	2
E4.1++	8	4		O4.5	646	20		X7-	52	1
E4.1+++	9	0		O4.6	112	9		X7	1	0
E4.2-	12	1		O4.6-	386	6		X7+	1378	91
E4.2+	225	20		O4.6--	3	0		X7++	4	1
E5-	117	1		O4.6+	1140	72		X8+	307	8
E5+	21	0		O4.6++	2	0		X8+++	1	0
E6-	126	2		P1	562	10		X9.1-	26	0
E6+	29	0		Q1.1	209	2		X9.1	9	0
F1	28451	1845		Q1.2	1206	8		X9.1+	205	0
F1-	33	11		Q1.2-	2	0		X9.1++	1	0
F2	4464	151		Q1.3	237	0		X9.2-	91	1
F2++	15	0		Q2.1	967	26		X9.2	7	0
F2+++	2	0		Q2.1-	7	0		X9.2+	364	3
F3	300	10		Q2.2	1530	19		X9.2++	7	0
F4	330	5		Q2.2-	1	0		Y1	228	1
G1.1	378	1		Q2.2+++	1	0		Y2	425	0
G1.1-	2	0		Q3	372	11		Z1	6100	65
G1.2	89	4		Q4	233	0		Z2	3530	66

USAS tag	Web	elicited		USAS tag	Web	elicited		USAS tag	Web	elicited
G2.1-	154	12		Q4.1	460	5		Z3	2445	58
G2.1	226	3		Q4.2	252	2		Z4	995	35
G2.1+	9	0		Q4.3	380	12		Z5	79990	1822
G2.2-	182	6		S1.1.1	516	4		Z6	1454	55
G2.2	11	0		S1.1.2+	144	2		Z7	528	29
G2.2+	228	10		S1.1.3-	13	1		Z7-	1	0
G3	317	0		S1.1.3+	176	0		Z8	15551	942
H1	695	11		S1.1.3+++	2	0		Z99	11612	101
H2	581	4		S1.1.4+	96	2		Z99+++	2	0
H3	61	1		S1.2	51	2				

Table F. Step 3: Wine: Correlation table using semantic word list (Step 3)

USAS tag	Web	elicited		USAS tag	Web	elicited		USAS tag	Web	elicited
A1.1.1	2636	193		G3-	4	0		S1.2	53	1
A1.1.1-	1	0		G3	3	0		S1.2.1-	31	0
A1.1.2	238	0		H1	779	7		S1.2.1+	188	3
A1.1.2-	1	9		H2	889	11		S1.2.2-	31	0
A1.2	16	0		H3	93	2		S1.2.2+	15	0
A1.2-	3	0		H4	655	14		S1.2.3-	13	0
A1.2.4-	1	0		H4-	3	0		S1.2.3+	52	9
A1.2+	126	0		H5	525	4		S1.2.3+++	1	0
A1.3-	9	0		I1	527	5		S1.2.4-	17	1
A1.3--	1	0		I1.1	402	4		S1.2.4	1	0
A1.3+	88	2		I1.1-	49	1		S1.2.4+	142	4
A1.4	153	2		I1.1+	159	3		S1.2.5-	25	1
A1.4-	7	0		I1.1++	3	0		S1.2.5--	2	0
A1.4+	16	0		I1.1+++	23	0		S1.2.5+	149	6
A1.5	1	0		I1.2	234	2		S1.2.5++	7	2
A1.5.1	457	30		I1.2-	2	0		S1.2.5+++	2	0
A1.5.1-	2	1		I1.2+	1	0		S1.2.6-	41	0
A1.5.1+	4	0		I1.3--	12	8		S1.2.6---	1	0
A1.5.2-	3	0		I1.3-	187	35		S1.2.6+	36	1
A1.5.2	3	0		I1.3---	4	0		S2	815	61
A1.5.2+	48	1		I1.3	577	0		S2.1	191	19
A1.6	27	0		I1.3+	48	41		S2.2	355	15
A1.7-	128	2		I1.3++	1	0		S3.1	372	40
A1.7+	145	4		I2.1	485	3		S3.1-	2	0
A1.7+++	1	9		I2.1-	1	0		S3.2	259	12
A1.8-	45	1		I2.2	1739	62		S3.2+	7	0
A1.8+	889	11		I3.1	382	7		S4	712	52
A1.9	53	1		I3.1-	36	0		S4-	2	0
A1.9-	1	0		I3.2-	10	0		S5-	178	1
A10-	269	3		I3.2	48	1		S5	3	0
A10+	844	8		I3.2+	40	2		S5+	1089	10
A11.1-	41	0		I4	144	1		S5+++	4	0
A11.1+	467	2		K1	730	22		S6-	187	2
A11.1++	51	0		K2	597	9		S6+	679	49

USAS tag	Web	elicited		USAS tag	Web	elicited		USAS tag	Web	elicited
A11.1+++	24	0		K3	64	0		S7.1-	92	1
A11.2-	16	0		K4	264	4		S7.1--	9	0
A11.2+	90	0		K5	72	0		S7.1	20	0
A12-	188	9		K5.1	491	1		S7.1-	2	0
A12+	145	3		K5.2	92	2		S7.1+	1376	11
A12++	9	0		K6	51	0		S7.1++	5	0
A12+++	3	0		L1	139	0		S7.2-	13	0
A13	96	1		L1-	184	1		S7.2+	160	0
A13.1	191	4		L1+	140	3		S7.3	23	0
A13.2	436	4		L2	1080	19		S7.3+	25	0
A13.3	1273	85		L3	1425	8		S7.4-	47	0
A13.4	261	12		L3-	1	0		S7.4	2	0
A13.5	124	11		M1	2287	57		S7.4+	242	6
A13.6	224	3		M2	1358	23		S8-	78	1
A13.7	97	2		M3	585	9		S8+	1271	18
A14	635	14		M4	448	6		S9	1734	86
A15-	73	2		M5	133	1		S9-	1	0
A15	2	0		M6	3700	61		T1	292	31
A15+	40	1		M7	1992	28		T1.1	65	2
A15++	2	0		M8	213	3		T1.1.1	425	28
A15+++	1	0		N1	1595	34		T1.1.2	674	16
A2.1-	35	0		N2	61	1		T1.1.3	1081	15
A2.1	1	0		N3	22	1		T1.2	502	16
A2.1+	696	41		N3.1	66	0		T1.3	3126	59
A2.2	1123	67		N3.2-	378	8		T1.3-	64	1
A2.2-	3	1		N3.2--	13	0		T1.3+	104	3
A2.2+	4	0		N3.2---	2	0		T1.3++	18	1
A3	15	0		N3.2	101	3		T2-	369	8
A3-	6	0		N3.2+	502	12		T2+	606	6
A3+	5414	604		N3.2++	31	1		T2++	356	5
A4.1	875	25		N3.2+++	107	2		T2+++	26	0
A4.2-	30	1		N3.3---	4	0		T3---	90	1
A4.2--	1	0		N3.3-	94	1		T3-	974	0
A4.2+	513	14		N3.3	420	1		T3--	63	0
A5.1---	19	2		N3.3--	6	0		T3	232	17
A5.1--	11	2		N3.3+	72	0		T3-	9	0
A5.1	284	6		N3.3++	4	0		T3+	320	11
A5.1-	74	30		N3.4	454	0		T3++	108	5
A5.1+	2050	158		N3.4-	1	0		T3+++	44	0
A5.1++	169	13		N3.4+	1	0		T4--	64	1
A5.1+++	794	26		N3.5	218	2		T4-	60	0
A5.1+++++	2	0		N3.5--	1	0		T4	1	0
A5.2-	75	0		N3.5-	2	0		T4+	71	1
A5.2+	243	2		N3.5+	26	2		W1	527	9
A5.2+++	1			N3.6	69	0		W2	212	4
A5.3-	40	2		N3.6-	1	0		W2-	59	3
A5.3+	91	0		N3.6+	1	0		W2--	3	0
A5.4-	57	0		N3.7-	82	1		W3	1016	10

USAS tag	Web	elicited		USAS tag	Web	elicited		USAS tag	Web	elicited
A5.4+	99	5		N3.7--	2	0		W4	146	2
A6	9	0		N3.7---	4	0		W5	72	0
A6.1-	924	32		N3.7	122	3		X1	28	0
A6.1	32	0		N3.7+	278	3		X2	43	5
A6.1+	372	32		N3.7++	28	0		X2.1	697	137
A6.1+++	145	3		N3.7+++	24	1		X2.1+	2	0
A6.2-	202	6		N3.8-	42	1		X2.2--	2	0
A6.2--	2	1		N3.8	65	0		X2.2-	75	2
A6.2+	537	24		N3.8+	153	5		X2.2	21	0
A6.3+	290	0		N3.8++	2	0		X2.2+	881	15
A6.3++	2	0		N3.8+++	5	0		X2.2++	4	0
A7	102	3		N4	1301	12		X2.2+++	11	0
A7-	49	1		N5---	201	3		X2.3+	97	2
A7+	1838	201		N5	1510	53		X2.4	290	0
A7++	1	0		N5-	432	12		X2.5-	27	2
A7+++	22	0		N5--	39	1		X2.5+	84	1
A8	241	0		N5.1-	383	3		X2.6-	61	0
A8+	2	0		N5.1	7	0		X2.6	2	0
A9-	1416	47		N5.1+	1704	42		X2.6+	152	1
A9	39	0		N5.1++	9	0		X3	29	0
A9+	2325	110		N5.1+++	10	0		X3-	2	0
B1	1445	68		N5.2+	193	43		X3.1	1023	85
B2-	305	62		N5+	1236	60		X3.1+	117	12
B2	38	6		N5++	1365	18		X3.2	257	0
B2+	64	5		N5+++	137	9		X3.2-	39	0
B2++	2	0		N6-	133	5		X3.2--	2	0
B3	271	6		N6	87	20		X3.2+	16	0
B4	271	0		N6+	364	39		X3.2++	1	0
B5	528	4		N6+++	83	11		X3.3	48	1
B5-	12	0		O1	357	6		X3.4	414	13
C1	555	7		O1.1	1223	44		X3.4-	14	1
E1	54	7		O1.2-	153	3		X3.4+	11	1
E1-	1	0		O1.2	665	19		X3.5	117	12
E2-	25	10		O1.3	115	1		X4.1	334	6
E2	3	0		O1.3-	1	0		X4.2	429	6
E2+	745	169		O2	3230	80		X5.1-	24	1
E2++	56	28		O3	83	0		X5.1+	98	0
E2+++	74	23		O4.1	853	8		X5.1++	3	0
E3-	276	5		O4.2-	87	22		X5.2-	79	2
E3--	1	0		O4.2	20	0		X5.2+	560	0
E3	1	0		O4.2+	1043	82		X5.2++	9	0
E3+	251	28		O4.2++	17	0		X6	19	0
E4.1-	132	5		O4.2+++	5	3		X6-	3	0
E4.1--	1	0		O4.3	2244	306		X6+	99	1
E4.1+	447	55		O4.4	338	4		X7-	37	0
E4.1++	2	0		O4.4+++	1	0		X7	2	0
E4.1+++	1	0		O4.5	312	0		X7.2+	3	0
E4.2-	19	1		O4.6-	223	15		X7+	1293	49

USAS tag	Web	elicited		USAS tag	Web	elicited		USAS tag	Web	elicited
E4.2+	131	8		O4.6--	6	0		X7++	2	0
E5-	87	2		O4.6	35	6		X8+	222	0
E5+	15	0		O4.6+	381	13		X8++	1	0
E5++	1	0		O4.6++	3	0		X8+++	1	0
E6-	100	2		P1	672	9		X9.1	10	0
E6+	63	4		Q1.1	280	3		X9.1-	17	0
F1	10137	282		Q1.2	1327	14		X9.1+	277	2
F1-	22	0		Q1.3	169	2		X9.1++	1	0
F2	13588	1869		Q2.1	1021	19		X9.1+++	2	0
F2-	2	0		Q2.1-	1	0		X9.2-	71	1
F2++	68	79		Q2.2	1895	19		X9.2	11	0
F2+++	12	2		Q2.2-	1	0		X9.2+	344	6
F3	193	2		Q3	427	12		X9.2+++	1	0
F3-	1	0		Q4	162	0		Y1	197	1
F4	648	11		Q4.1	420	3		Y2	300	0
G1.1	653	0		Q4.2	281	0		Z1	5578	106
G1.1-	3	0		Q4.3	251	4		Z2	4630	143
G1.2	120	0		S1.1.1	672	17		Z3	1697	8
G2.1-	117	5		S1.1.1-	1	0		Z4	782	18
G2.1	363	3		S1.1.2-	1	0		Z5	85787	2233
G2.1+	14	0		S1.1.2+	116	6		Z6	1414	72
G2.2-	205	3		S1.1.3-	17	0		Z7	413	25
G2.2	11	0		S1.1.3+	246	7		Z8	14627	876
G2.2+	162	1		S1.1.3+++	4	0		Z99	10229	136
G2.2++	1	0		S1.1.4	1	0				
G3	233	1		S1.1.4+	85	0				

The golden section in texts

Arjuna Tuzzi*, Ioan-Iovitz Popescu, Gabriel Altmann

*Dip. di Sociologia, Università di Padova, Via Cesarotti 12, 35123 Padova, Italia

Abstract

The golden section is a well known phenomenon observed in nature, arts and sciences, documented with an enormous number of publications. Here we shall try to show its presence in the rank-frequency distribution of words in natural texts with emphasis on Italian using the End-of-Year Speeches of the Presidents of the Italian Republic.

Keywords: Golden section, Italian texts

Introduction

In Euclid of Alessandria's definition a straight line is said to have been cut in extreme and mean ratio when, as the whole line is to the greater segment, so is the greater to the lesser. This ratio later became known as the "golden section", also called "golden ratio", "golden mean", "Divine proportion", etc. In recent years Mario Livio (2002a) discussed in depth the appearances of the golden section in nature, arts, psychology, etc. and affirmed that the golden section has inspired scholars of many disciplines like no other number in the history of mathematics. In different domains of nature (botanics, human body, self-organized growth etc.), in visual arts, music, architecture, some phenomena of poetry and in some sciences the golden section seems to be often a rather manifest phenomenon that can be stated by quite simple measurement. In mathematics it can be derived. But in textology it is rather latent and can be arrived at indirectly using some properties of texts.

In some previous studies (cf. Popescu & Altmann 2007; 2009) it has been observed that a special relationship of some properties of ranked word-form frequencies yields the golden section. Though evidence has been collected from 176 texts in 20 languages and 253 texts of 26 German authors, the necessity of further corroboration is an ever lasting stimulus for further research. To this end we used 60 Italian texts, namely the End-of-Year Speeches of the ten Presidents of the Italian Republic (1949-2008: Luigi Einaudi, Giovanni Gronchi, Antonio Segni, Giuseppe Saragat, Giovanni Leone, Sandro Pertini, Francesco Cossiga, Oscar Luigi Scalfaro, Carlo Azeglio Ciampi and Giorgio Napolitano) which, at least concerning their aim, display a kind of homogeneity. Besides, we present graphically also the results obtained from the translations of the same text in Slavic languages (Kelih, 2009). The corpus is interesting for several reasons (Pauli and Tuzzi, 2009; Cortelazzo and Tuzzi, 2007): in Italy the President of Republic is the most authoritative Office, the speeches are rich in information about the recent history, and the End-of-Year Speech is an important "civil ritual" (delivered first on the radio, then, since 1956, at television, and in more recent years simultaneously broadcasted on television by the main public and private national networks, cf. Zotti Minici, 2007).

The appearance of the golden section in other domains of culture is a well known fact described in many books and articles (for references see e.g. en.wikipedia.org/wiki/Golden_ratio or http://www.goldenmuseum.com/1801Refer_engl.html), but in texts which are created under stochastic regimes its existence can, in best case, be rather a matter of convergence – if it exists at all. It cannot be deduced as a certain ratio, it can only be observed as a convergence to the irrational golden number $\phi = (1+\sqrt{5})/2 \cong 1.6180$. But even if we find such a convergence, our assumption need not be at any cost correct because in nature and in mathematics there are a number of constants which can be found in some textual phenomena. Nevertheless, we try to find the ways to this number and test its existence on texts of different kind.

Data and method

In the study of rank-frequencies of word-forms there are several ways to characterize the form of this decreasing sequence of numbers. The first consists in the subconscious control of word repetitions by the writer. The writer is assumed "to sit" somewhere in the mid of the arising rank-frequency sequence and to control the proportionate increase of frequent words (usually synsemantics) and the slow increase of autosemantic words. The fixed point at which he "sits" is called the *h*-point. This index was first suggested by Hirsch (2005) as a tool for quantifying the scientists' publication productivity and then introduced in linguistics by Popescu (2006).

In a rank-frequency distribution of word forms the *h*-point is computed according to the formula

$$(1) \quad h = \begin{cases} r & \text{if there is an } r = f(r) \\ \frac{f(r_i)r_j - f(r_j)r_i}{r_j - r_i + f_i - f_j} & \text{if there is no } r = f(r) \end{cases}$$

where r is the rank and $f(r)$ is the frequency of the word at rank r . If there is no $r = f(r)$, one takes two neighbouring ranks for which $r_i < f(r_i)$ and $r_j > f(r_j)$. In case that $r_{max} < f(r_{max})$ one transforms the frequency sequence in $f^*(r) = f(r) - f(r_{max}) + 1$. To the left of the *h* point the writer "sees" the synsemantics, to the right the autosemantics, though, as well known, this boundary is fuzzy.

The four cardinal points of any rank-frequency distribution are

1. origin $O(1,1)$ (and not $O(0,0)$)
2. top $P_1(1, f_1)$ (f_1 being the greatest frequency)
3. end $P_2(V,1)$ (V = vocabulary)
4. *h*-point $H(h,h)$.

The writer's view is defined as the angle (alpha) subtended between the vectors

HP_1 of Cartesian components $a_x = -(h - 1)$, $a_y = f_1 - h$
 HP_2 of Cartesian components $b_x = V - h$, $b_y = -(h - 1)$.

Inserting these components in the general expression

$$(2) \quad \cos \alpha = \frac{a_x b_x + a_y b_y}{[(a_x^2 + a_y^2)^{1/2}][(b_x^2 + b_y^2)^{1/2}]}$$

we obtain the exact formula

$$(3) \quad \cos \alpha = \frac{-[(h-1)(f_1 - h) + (h-1)(V - h)]}{[(h-1)^2 + (f_1 - h)^2]^{1/2}[(h-1)^2 + (V - h)^2]^{1/2}}$$

This leads to a slight improvement of α -values, of the percent order, as compared to those obtained with our original formula (deduced with the origin $O(0,0)$ instead of $O(1, 1)$). Actually, the original writer's view formula (Popescu, Altmann 2007) differs from Eq. (3) by merely replacing $(h - 1)$ through h so that both expressions give practically the same results for $h \gg 1$.

Finally, it should be pointed out that for the limiting case of $H(1,1)$, that is for $h = 1$, $\cos \alpha$ is zero and, correspondingly, $\alpha = \pi/2$ radians = 1.57079633... radians. However, in practical texts this "orthogonality" is never attained, but is apparently replaced by a limit close to the golden section 1.6180... As a matter of fact, in the limit of large texts it appears that the h -point makes the difference between $\pi/2$ and the golden number.

Results

In order to illustrate the approximation of writer's view to the golden section we show some numerical results in Table 1 and the trend in terms of text size in Figures 1 to 4 below. Generally, the α radians approach the golden section with increasing text length. This fact can be interpreted under the assumption that in short text the writer begins to search subconsciously for a kind of equilibrium but can achieve it only if the text develops. According to Orlov (1982), frequencies are distributed in agreement with the "planned" length of the text, called now Zipf-Orlov size. The "planned" flow of information - which is conscious and should bear some features of originality - is at variance with the general subconscious mechanism leading to an equilibrium expressed e.g. by the golden section. It can be arrived at only if the text increases and the writer loosens his original restricted aim. The mechanism of writing settles and the text gets proportions which are present in all human artistic activity.

Table 1
Writer's view of 60 End-of-Year speeches of Italian Presidents (1949 - 2008)

Text	N	V	$f(1)$	h	$\cos \alpha$	α rad
1949Einaudi	194	140	10	5.00	-0.6475	2.2752
1950Einaudi	150	105	9	4.00	-0.5397	2.1409
1951Einaudi	230	169	9	5.00	-0.7241	2.3806
1952Einaudi	179	145	7	4.00	-0.7220	2.3775
1953Einaudi	190	143	8	4.00	-0.6171	2.2359
1954Einaudi	260	181	12	5.00	-0.5157	2.1127
1955Gronchi	388	248	16	6.66	-0.5382	2.1391
1956Gronchi	665	374	29	8.00	-0.3343	1.9117
1957Gronchi	1130	549	65	12.00	-0.2232	1.7959
1958Gronchi	886	460	41	11.00	-0.3373	1.9148

1959Gronchi	697	388	33	9.00	-0.3362	1.9137
1960Gronchi	804	434	41	10.00	-0.2991	1.8746
1961Gronchi	1252	622	67	13.00	-0.2361	1.8092
1962Segni	738	381	35	10.00	-0.3614	1.9406
1963Segni	1057	527	46	11.66	-0.3162	1.8925
1964Saragat	465	278	21	8.00	-0.4968	2.0907
1965Saragat	1052	510	52	11.66	-0.2761	1.8505
1966Saragat	1200	597	44	12.50	-0.3614	1.9405
1967Saragat	1056	526	51	11.00	-0.2613	1.8352
1968Saragat	1173	562	56	13.00	-0.2898	1.8648
1969Saragat	1583	692	86	15.00	-0.2137	1.7862
1970Saragat	1929	812	85	16.50	-0.2397	1.8128
1971Leone	262	168	12	5.00	-0.5173	2.1145
1972Leone	767	394	32	9.50	-0.3740	1.9541
1973Leone	1250	616	67	12.00	-0.2139	1.7864
1974Leone	801	426	32	9.00	-0.3466	1.9247
1975Leone	1328	632	63	13.00	-0.2522	1.8257
1976Leone	1366	649	52	13.00	-0.3121	1.8882
1977Leone	1604	717	80	14.00	-0.2114	1.7838
1978Pertini	1492	603	53	14.33	-0.3472	1.9254
1979Pertini	2311	800	70	18.00	-0.3313	1.9085
1980Pertini	1360	535	50	13.75	-0.3548	1.9335
1981Pertini	2819	911	96	20.00	-0.2632	1.8371
1982Pertini	2486	854	90	19.00	-0.2666	1.8406
1983Pertini	3746	1149	118	23.66	-0.2531	1.8267
1984Pertini	1340	514	42	13.66	-0.4308	2.0162
1985Cossiga	2359	859	118	17.00	-0.1752	1.7469
1986Cossiga	1348	561	65	14.00	-0.2700	1.8441
1987Cossiga	2092	904	109	15.00	-0.1629	1.7344
1988Cossiga	2384	875	123	19.00	-0.1912	1.7632
1989Cossiga	1912	778	85	17.00	-0.2495	1.8229
1990Cossiga	3345	1222	155	20.00	-0.1550	1.7264
1991Cossiga	418	241	22	7.00	-0.3951	1.9769
1992Scalfaro	2774	978	118	17.50	-0.1789	1.7507
1993Scalfaro	2942	1074	129	18.60	-0.1739	1.7456
1994Scalfaro	3606	1190	171	21.00	-0.1491	1.7205
1995Scalfaro	4233	1341	180	22.66	-0.1526	1.7240
1996Scalfaro	2085	866	88	16.00	-0.2212	1.7938
1997Scalfaro	5012	1405	167	27.50	-0.2055	1.7778
1998Scalfaro	3995	1175	137	23.50	-0.2136	1.7860
1999Ciampi	1941	831	66	16.50	-0.3169	1.8933
2000Ciampi	1844	822	70	16.00	-0.2855	1.8604
2001Ciampi	2098	898	89	18.00	-0.2516	1.8251
2002Ciampi	2129	909	96	17.00	-0.2160	1.7886
2003Ciampi	1565	718	63	14.00	-0.2742	1.8486
2004Ciampi	1807	812	76	15.00	-0.2408	1.8140

2005Ciampi	1193	538	54	12.66	-0.2927	1.8679
2006Napolitano	2204	929	125	16.50	-0.1582	1.7297
2007Napolitano	1792	793	101	16.00	-0.1928	1.7648
2008Napolitano	1713	775	75	15.00	-0.2451	1.8184

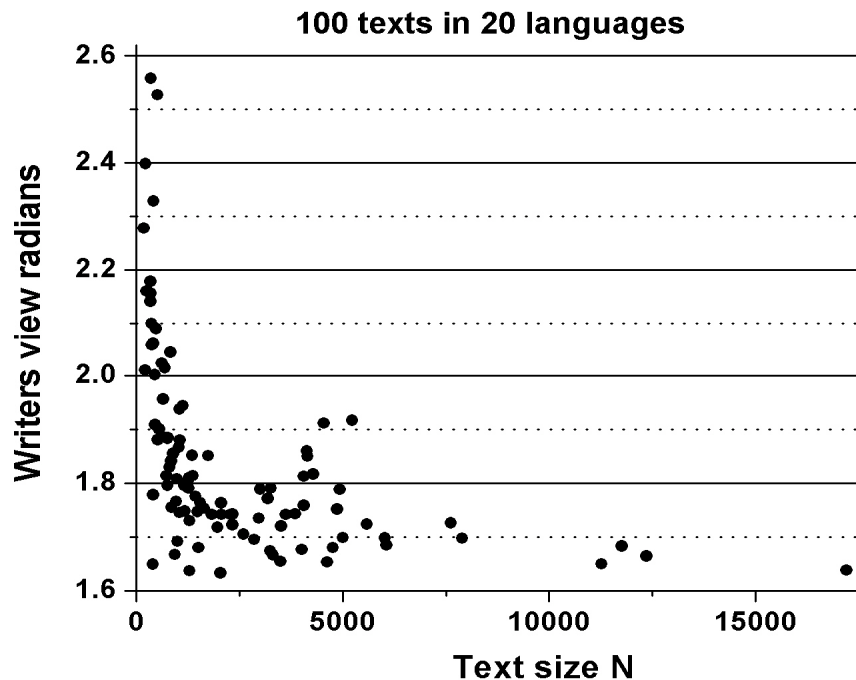


Fig. 1. α radians in 100 texts in 20 languages (data from Popescu, Altmann 2009)

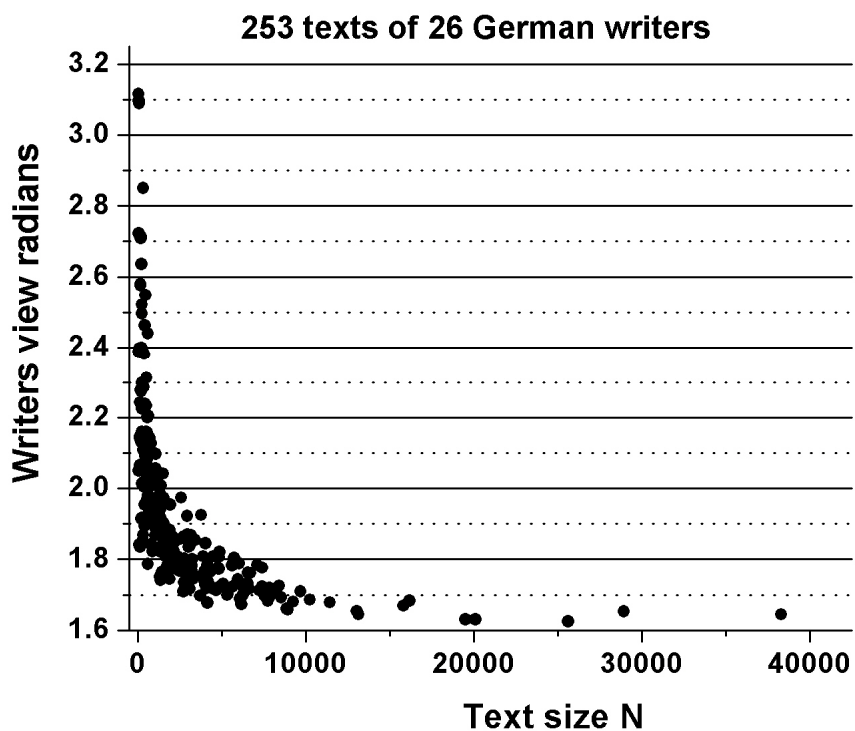


Figure 2. α radians in 253 German texts (data from Popescu, Altmann 2009)



Fig. 3. α radians in 120 Slavic texts
(data from Emmerich Kelih's project on Slavic languages 2009)

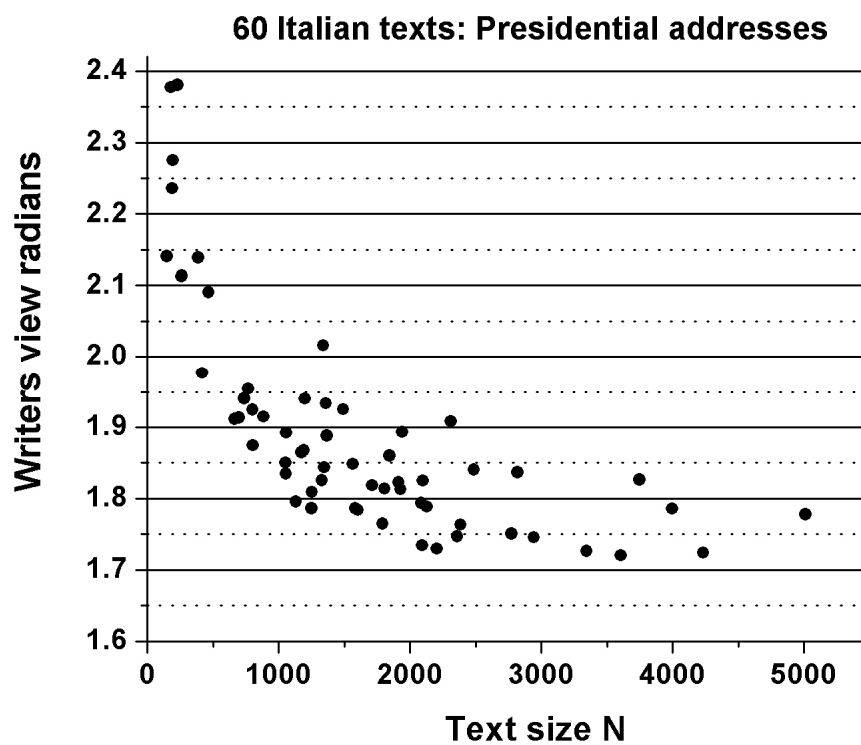


Figure 4. α radians in 60 Italian texts: Presidential addresses
(data from Tuzzi, Popescu, Altmann 2009)

In Slavic texts which are translations of ten chapters of a Russian novel by Ostrovskij in eleven languages, the great deviations are caused rather by the fact that the translators did not have the possibility of creating the text spontaneously but were pressed in an a priori given form which did not allow to make decisions either concerning text length or information flow. Nevertheless, the convergence is evident.

In Italian texts which are rather short, the convergence is obvious but it does not approach the golden section sufficiently, a fact caused by the small text sizes.

The existence of the mechanism, even if working differently in different languages can, nevertheless, be considered as given.

Let us approach the problem from a different point of view. The appearance of the golden section can be demonstrated also using the arc length joining the greatest frequency f_1 with the greatest rank $V = r_{max} = R$. According to Hirsch (2005) text length N is associated with the h -point in form $N = ah^2$ but since the origin of the rank-frequency curve is $O(1,1)$ [and not $O(0,0)$] Popescu and Altmann (2009) defined a slightly differing association $N = b(h-1)^2$ and established an indicator which is characteristic for different language phenomena, namely

$$(4) \quad p = \frac{L_{max} - L}{h - 1}$$

where L is the arc length computed as a sum of Euclidean distances between the individual ordered frequencies:

$$(5) \quad L = \sum_{r=1}^{R-1} [(f(r) - f(r+1))^2 + 1]^{1/2}$$

and L_{max} is the maximal possible arc length given as

$$(6) \quad L_{max} = R - 1 + f(1) - 1.$$

If $f(R)$ is not 1, the formula must be slightly modified. Since h is a function of text size N , one can use another indicator corresponding to (4), namely

$$(7) \quad q = \frac{L_{max} - L}{N^{1/2}}.$$

Combining these two expressions, we obtain

$$(8) \quad p + q = (L_{max} - L) \left(\frac{1}{\sqrt{N}} + \frac{1}{h-1} \right)$$

which, strangely enough, converges to the golden section. This convergence can preliminarily not be demonstrated mathematically but we again show some cases from different languages and texts in order to at least register this fact.

Again we firstly illustrate some numerical data in Table 2 and the trend of p , q and $p + q$ in terms of text size in Figures 5 to 8 below.

Table 2
The (p,q) indicator pair of 60 End-of-Year speeches of Italian Presidents (1949 - 2008)

Text	N	V	$f(1)$	h	L	L_{\max}	p	q
1949Einaudi	194	140	10	5.00	143.54	148	1.1142	0.3200
1950Einaudi	150	105	9	4.00	108.78	112	1.0733	0.2629
1951Einaudi	230	169	9	5.00	172.23	176	0.9417	0.2484
1952Einaudi	179	145	7	4.00	146.89	150	1.0357	0.2322
1953Einaudi	190	143	8	4.00	145.82	149	1.0603	0.2308
1954Einaudi	260	181	12	5.00	186.29	191	1.1772	0.2920
1955Gronchi	388	248	16	6.66	255.36	262	1.1739	0.3373
1956Gronchi	665	374	29	8.00	392.75	401	1.1782	0.3198
1957Gronchi	1130	549	65	12.00	599.38	612	1.1476	0.3755
1958Gronchi	886	460	41	11.00	488.04	499	1.0956	0.3681
1959Gronchi	697	388	33	9.00	409.82	419	1.1475	0.3477
1960Gronchi	804	434	41	10.00	462.23	473	1.1969	0.3799
1961Gronchi	1252	622	67	13.00	674.05	687	1.0789	0.3659
1962Segni	738	381	35	10.00	404.01	414	1.1101	0.3678
1963Segni	1057	527	46	11.66	559.52	571	1.0771	0.3532
1964Saragat	465	278	21	8.00	289.03	297	1.1383	0.3695
1965Saragat	1052	510	52	11.66	547.78	560	1.1466	0.3768
1966Saragat	1200	597	44	12.50	624.77	639	1.2376	0.4109
1967Saragat	1056	526	51	11.00	562.98	575	1.2019	0.3699
1968Saragat	1173	562	56	13.00	602.83	616	1.0978	0.3847
1969Saragat	1583	692	86	15.00	759.82	776	1.1556	0.4066
1970Saragat	1929	812	85	16.50	877.58	895	1.1242	0.3967
1971Leone	262	168	12	5.00	173.02	178	1.2443	0.3075
1972Leone	767	394	32	9.50	414.71	424	1.0932	0.3355
1973Leone	1250	616	67	12.00	669.22	681	1.0710	0.3332
1974Leone	801	426	32	9.00	445.78	456	1.2770	0.3610
1975Leone	1328	632	63	13.00	678.97	693	1.1688	0.3849
1976Leone	1366	649	52	13.00	685.16	699	1.1535	0.3745
1977Leone	1604	717	80	14.00	780.72	795	1.0982	0.3565
1978Pertini	1492	603	53	14.33	639.45	654	1.0918	0.3768
1979Pertini	2311	800	70	18.00	848.35	868	1.1558	0.4087
1980Pertini	1360	535	50	13.75	567.95	583	1.1800	0.4080
1981Pertini	2819	911	96	20.00	983.94	1005	1.1085	0.3967
1982Pertini	2486	854	90	19.00	921.74	942	1.1257	0.4064
1983Pertini	3746	1149	118	23.66	1236.65	1265	1.2513	0.4633
1984Pertini	1340	514	42	13.66	539.18	554	1.1704	0.4048
1985Cossiga	2359	859	118	17.00	955.75	975	1.2033	0.3964
1986Cossiga	1348	561	65	14.00	610.09	624	1.0699	0.3788
1987Cossiga	2092	904	109	15.00	993.76	1011	1.2312	0.3769
1988Cossiga	2384	875	123	19.00	976.91	996	1.0606	0.3910
1989Cossiga	1912	778	85	17.00	842.21	861	1.1742	0.4297
1990Cossiga	3345	1222	155	20.00	1351.79	1375	1.2214	0.4012

1991Cossiga	418	241	22	7.00	254.77	261	1.0384	0.3047
1992Scalfaro	2774	978	118	17.50	1072.80	1094	1.2848	0.4025
1993Scalfaro	2942	1074	129	18.60	1179.30	1201	1.2327	0.4000
1994Scalfaro	3606	1190	171	21.00	1333.26	1359	1.2869	0.4286
1995Scalfaro	4233	1341	180	22.66	1492.52	1519	1.2227	0.4071
1996Scalfaro	2085	866	88	16.00	934.04	952	1.1975	0.3934
1997Scalfaro	5012	1405	167	27.50	1538.44	1570	1.1908	0.4458
1998Scalfaro	3995	1175	137	23.50	1281.19	1310	1.2806	0.4559
1999Ciampi	1941	831	66	16.50	877.32	895	1.1405	0.4012
2000Ciampi	1844	822	70	16.00	871.20	890	1.2531	0.4377
2001Ciampi	2098	898	89	18.00	965.54	985	1.1446	0.4248
2002Ciampi	2129	909	96	17.00	984.94	1003	1.1287	0.3914
2003Ciampi	1565	718	63	14.00	763.50	779	1.1925	0.3919
2004Ciampi	1807	812	76	15.00	869.71	886	1.1639	0.3833
2005Ciampi	1193	538	54	12.66	576.22	590	1.1815	0.3989
2006Napolitano	2204	929	125	16.50	1033.53	1052	1.1918	0.3935
2007Napolitano	1792	793	101	16.00	874.57	892	1.1621	0.4118
2008Napolitano	1713	775	75	15.00	831.25	848	1.1961	0.4046

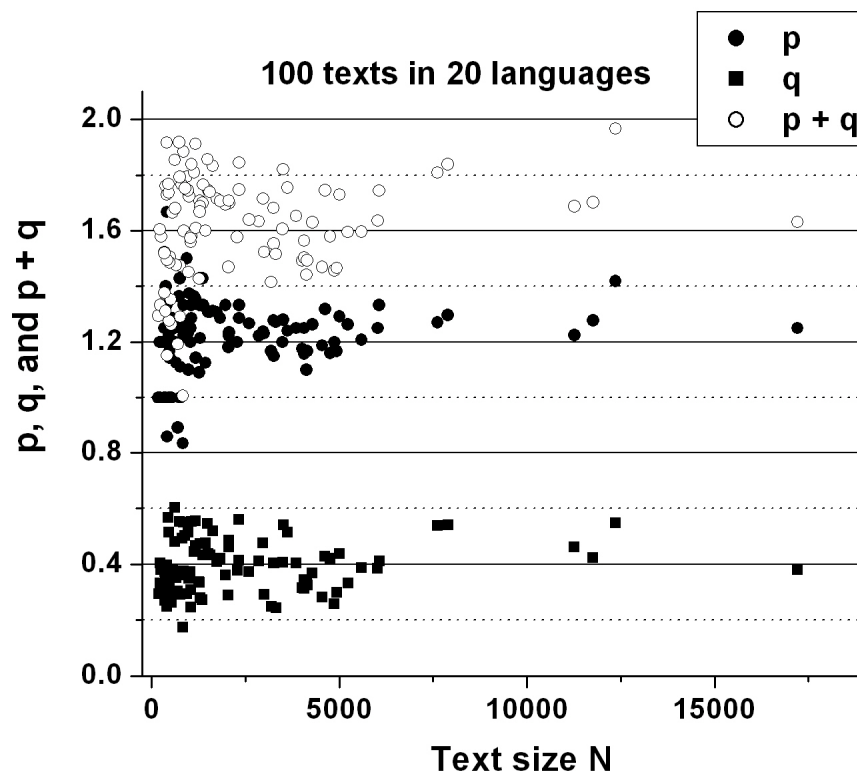


Figure 5. Showing p , q , and their sum in terms of the text size N in 100 texts in 20 languages (data from Popescu, Altmann 2009)

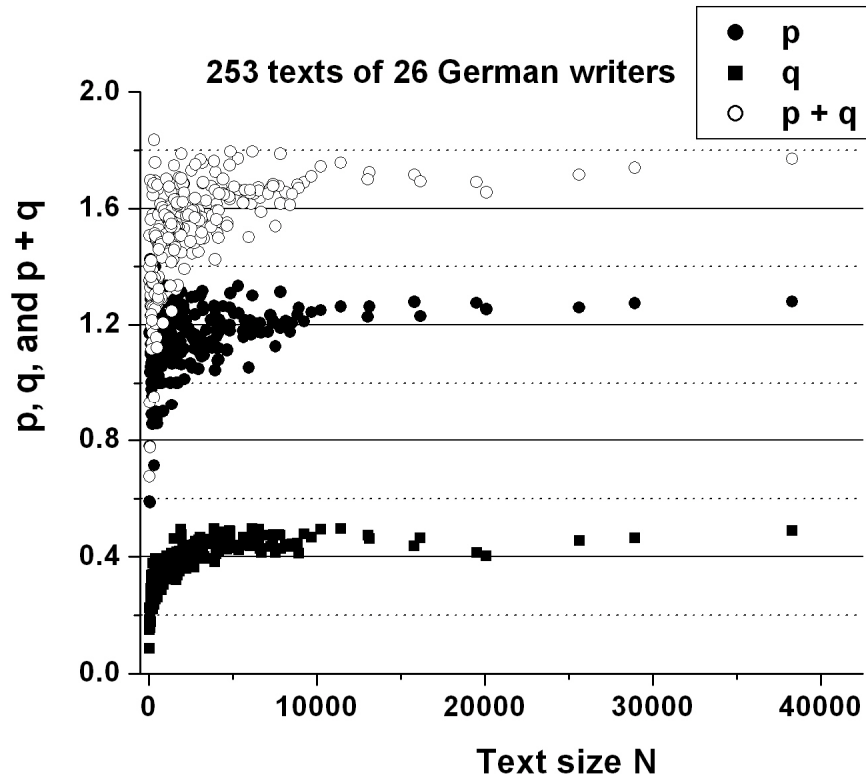


Figure 6. Indicators p , q , and their sum in terms of the text size N in 253 texts of 26 German authors (data from Popescu, Altmann 2009)



Figure 7. Indicators p , q , and their sum in terms of the text size N in 120 Slavic texts (data from Emmerich Kelih's project on Slavic languages 2009)

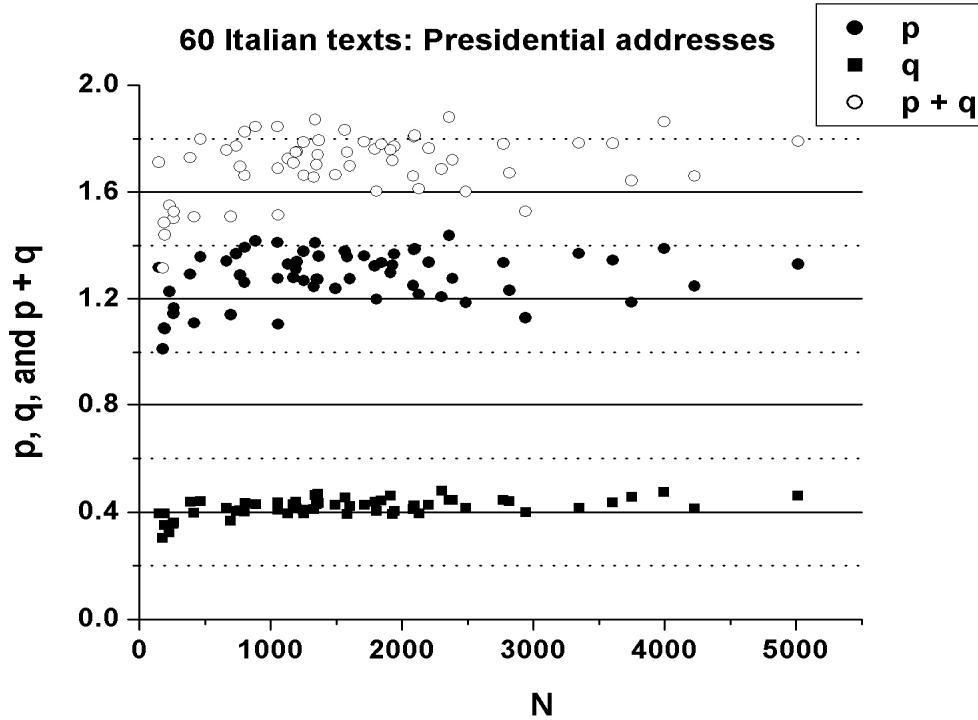


Figure 8. Indicators p , q , and their sum in terms of the text size N in 60 Italian texts: Presidential addresses (data from Tuzzi, Popescu, Altmann 2009)

Generally, as can be seen, q tends to about 0.4, p to about 1.2 and $p + q$ is dispersed around the golden section. For Italian literary texts the authors found $\bar{p} = 1.269$, $\bar{q} = 0.412$, yielding the sum 1.681. Now, for the addresses of the presidents we obtain the results presented in Table 2 and in Figure 8. Again, the mean $\bar{p} = 1.156$, $\bar{q} = 0.375$, the sum yielding 1.531, smaller than in Italian literary texts. This is, perhaps, caused by small text sizes. We assume that the tendency exists but except for text size no other causes of differences could be found. This is left to further research. Since the variances of all quantities defined here are known (cf. Popescu, Mačutek, Altmann 2009), setting up asymptotic tests for differences in p or q or $p + q$ between texts is always possible.

Discussion

A special relationship between text length, the arc length of the rank-frequency sequence of word forms and its h -point seems to converge to the golden section. Though we brought evidence from different languages and text types, further evidence is necessary. Since in empirical sciences every hypothesis gives rise to other ones or at least to new questions, one can ask (1) whether this phenomenon appears only in word frequencies or can be generalized to any other linguistic unit and (2) whether this is the only way to obtain the golden section in texts or are there perhaps other relationships leading to it. And last but not least, (3) what is the relationship of the textual golden section to its appearance in other forms of human activity? Is it generated by the same background mechanism or does it have another origin?

Although the existence of a most pleasing feature related to the golden section is open to criticism and although the golden section rules are only one of several possible theories explaining the involved concept of beauty, the appearances of the golden section in the arts fostered many attempts to investigate a potential relationship between the human perception

of beauty and the golden section (Livio, 2002b). The presumed association of the golden section with aesthetics was discussed even in the field of facial surgery by Marquardt (2002): he attempted to "quantify" the concept of beauty by developing the golden decagon mask (based upon the golden section) and some plastic surgeons use this model to enhance their patients' facial features.

Looking at the problem from another point of view we can ask whether the textual golden section evokes the same aesthetic impressions as its counterparts in art or music. Do all of them have a geometric background or is there a difference in perception? Is rhythm involved or is it a "feeling" for part-and-whole relation? Is it programmed genetically or does it depend on individual cultures? We shall never be able to give a "last" explanation of this phenomenon but a step by step treatment of individual problems can help us to capture ever more of its aspects.

Acknowledgments

The authors are greatly indebted to Emmerich Kelih for keeping at our disposal word form rank-frequency data of his project on Slavic languages.

References

- Cortelazzo M.A., Tuzzi A. (Eds.) (2007). *Messaggi dal Colle. I discorsi di fine anno dei presidenti della Repubblica*. Venice: Marsilio.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 102(46), 1656–1657.
- Kelih, E. (2009). Personal communication.
- Livio, M. (2002a). *The golden ratio: The story of phi, the extraordinary number of nature, art and beauty*. London: Headline.
- Livio, M. (2002b). The golden ratio and aesthetics. *Plus Magazine*, 22 (<http://plus.maths.org/>)
- Marquardt, S.R. (2002). Dr. Stephen R. Marquardt on the golden decagon and human facial beauty. *Journal of Clinical Orthodontics*, 36(6), 339-47.
- Orlov, J.K., Boroda, M.G., Nadarejšvili, I.Š. (1982). *Text, Sprache, Kunst. Quantitative Analysen*. Bochum: Brockmeyer.
- Pauli, F., Tuzzi, A. (2009). The End of Year Addresses of the Presidents of the Italian Republic (1948-2006): Discoursal similarities and differences. *Glottometrics* 18, 40-51.
- Popescu, I.-I. (2007). Text ranking by the weight of highly frequent words. In P. Grzybek & R. Köhler (Eds.), *Exact methods in the study of language and text*: 555-565. Berlin-New York: Mouton de Gruyter.
- Popescu, I.-I., Altmann G. (2007). Writer's view of text generation. *Glottometrics* 15, 71-81.
- Popescu, I.-I., & Altmann, G. (2009). A modified text indicator. In E. Kelih, V. Levickij, & G. Altmann (Eds.), *Problems of quantitative text analysis*. Cernivcy: RUTA.
- Popescu, I.-I., Mačutek, J., & Altmann, G. (2009). *Aspects of word frequencies*. Lüdenscheid: RAM-Verlag.
- Tuzzi, A., Popescu, I.-I., & Altmann, G. (2009). Zipf's laws in Italian texts. *Journal of Quantitative Linguistics* 16(4), 354-367.
- Zotti Minici, A. (2007). Dove guardano i presidenti a fine anno. In M. A. Cortelazzo & A. Tuzzi (Eds.), *Messaggi dal Colle. I discorsi di fine anno dei presidenti della Repubblica* (pp. 47-53). Venice: Marsilio.

Character functions as indicators of self states in life stories¹

Bernadette Péley*, János László

*University of Pécs, Institute of Psychology
6. Ifjúság útja, Pécs, Hungary H-7624

Abstract

Experiences of emotion regulation and self-experiences derive from early mother-child interaction. In later stages of life, these experiences regulate how individuals perceive and interpret their interactions with other people and manage their interpersonal relations. When reconstructing their significant life episodes, the psychological functions they implicitly assign to their partners reflect this perception and interpretation and thereby their self-state. The paper presents a computer program that identifies positive and negative character functions in life stories and a validity study which supports that the program measures self-states.

Keywords: self development, object relations, life stories, character functions, computerized content analysis

Introduction

One of the major assumptions of narrative psychology is that meaning and meaning construction have diagnostic as well as predictive significance for individual and group adaptation. Narrative has been conceived as biologically given but culturally mediated form of meaning construction. Although narrative psychology has been initiated as a hermeneutic enterprise (Ricoeur, 1984-1987; Sarbin, 1986; Gergen and Gergen, 1988 Polkinghorn, 1995; Bamberg, 2007), narrative contents and narrative forms provide ground for studying identity processes empirically, which are held basic for social adaptation.

In a series of studies, László and his research group (László et al. 2007; Ehmann et al. 2007; Hargitai et al. 2007; Pohárnok et al. 2007; Pólya et al. 2007; Vincze et al. 2007) have attempted to operationalise the links between narrative and psychological constructs by developing computer algorithms that are able to identify narrative categories such as narrative perspective, temporal organization, organizing spatial and emotional distance, self-reference, and negation.

The present study turns to another classic narrative category, i.e. characters' functions (see Propp, 1967). We assume that functions of life story characters are not shaped by chance;

¹ This research was supported by the OTKA T38387 grant and Bolyai Scholarship to the first author NKFP6-00074/2005 grant to the second author

*peley@btk.pte.hu

instead, they show patterns that are characteristic to the story teller. In a developmental perspective, life story can be related to the self- and object representations tracing back to the early mother-infant interactions (Péley, 2002; Péley and László, 2009). Representations of the self-with-others events serve as basis for self knowledge. Self knowledge is understood in two senses. On the one hand, it is conscious and reflected in stories. On the other hand, it is non-conscious knowledge, and it organizes our memories of ourselves and of our relationships. This latter kind of knowledge originates from the preverbal age. There are several observations, experiments and theories, which show that quality of early interpersonal relationships are decisive for integrity and coherence of the self (Stern, 1989, 1995; Gergely and Watson, 1996; Siegel, 2001; Beebe and Lachmann, 2002.).

Role of narrative in self-development

According to Stern (1989), different senses of self are developing in the first years of life. Core self, which consists of senses such as agency, coherence, continuity or affectivity, emerges in the second-third month. This self is non-reflexive and non-conscious.

Around the ninth month the subjective sense of self emerges. It is also non-reflexive and non-conscious. The infant experiences her own states of consciousness, which can be shared with others, who, on the other hand, also have their own subjective mental states. This new subjective self enables intersubjectivity and interactions based on intersubjectivity. Mental contents, (e.g., attentional focus, intentions or emotions) can be shared with others. A “verbal” sense of self evolves around the fifteenth-eighteenth month. It is already self-reflexive, enabling the objectivation of the self. Self-reflexivity shows in using personal pronouns or in the infant’s behavior in front of a mirror.

Each sense of self organizes subjective perspectives to the self. At the time of developmental shifts, new cognitive, affective, and motivational capacities emerge as consequences of maturation. Infants have to form a new perspective towards themselves at these stages, which organizes the new capacities. Senses of self are organizers of the subjective perspective in this way. Each new sense of self opens up new domains of experiences, but it does not incorporate, dissolve or eliminate the earlier ones. Senses of self live together. Narrative sense of self emerges at around the end of the second year. It partly derives from the new linguistic and conceptual capacities. These new capacities not only enable but also force the child to re-organizing her subjective perspective in narrative form. Senses of agency, coherence, continuity, affectivity, intersubjectivity, and self-reflection are re-organized already in narratives. Crib monologues, which can be observed at around age two, have the function of consolidating the narrative self (Bruner and Lucariello, 1989). It is this self which will serve as the basis for the whole life when somebody accounts for her life to herself or to others.

When the narrative self unfolds, the child passes the border between the pasts which can be versus cannot be reconstructed. Narrative, however, does not eliminate the process of giving meaning to life events, i.e. application of subjective perspective. Rather, this meaning construction is performed by the narrative. Recalling and telling life episodes indicates how the person gives meaning to the events, what kind of organizing principles are applied, and the nature of the constructive processes which take part in the reconstruction.

Anything that happens to us is appropriated by giving meaning to it. In this way it adds to self-organization and to self knowledge. It is not the event itself that is preserved, but its meaning for the self, for the others, for the relationships. It follows that recollection of a childhood memory and telling a story about it is not a copy of the event. Rather, it is reconstructive act in which the story teller constructs a story out of her memories in the “hic et nunc” situation of the story telling.

Similar thoughts were outlined by Barclay and Smith (1992) who assigned an eminent role to early object relations in the organization and recall of autobiographical memories. Relying on Winnicott's developmental theory (Winnicott, 1990), they stress the regulating function of these memories by allocating narrated, shared memories among "transitory objects". As they put it, "Memories, and autobiographical memories especially, are precisely such transitional phenomena – symbols of the deep templates of caregiving we come to rely upon, again and again, to serve our needs for emotional support and responsiveness. We release such constructions into the space between ourselves and significant others as a way of recapitulating the caregiving and "structure-building" that occurred in our pasts. No less than a child will change its story to suit its feelings, we also adapt our memories as our needs change or as the needs of our partners change. Indeed, reconstructed memories in this sense become a kind of currency of social life, especially for the purchase of intimacy. It is only in the presence of trusted others that we really surrender our illusions, and indeed bring them under another's control. Serving a significant other's needs, likewise, we adapt our memories in natural responsiveness to their emotional states." (Barclay and Smith, 1992, p.89.) Autobiographical memories and life stories presenting these memories thus are used to regulating our emotional life.

These ideas entail a further assumption: life stories of well adapted personalities will differ from those of maladaptive personalities with respect to emotional regulation. It is an empirical question how these differences will show in life stories?

Characters' functions and plot units

In a brilliant analysis of Russian fairy tales, Propp (1968) concluded that, despite the great variety of their plots, these stories combined a very limited number of plot units. He also discovered that combination of plot units was governed by grammar-like rules. Characters' actions may differ on the surface, however they can still be generalized according to their intention. These generalized actions are plot units. For instance, there are several ways to hurt or damage another people, e.g., beat, kill, rob, etc. Propp himself counted twenty different forms in the fairy tale corpus. No matter whether the villain beats, kills, or robs the hero or his associate, the action evolves the tale in a certain direction. Given that Propp studied plot units with respect to their functions in the construction of the tales and plot units are performed by the characters, he named them characters' functions or simply functions. Propp also observed that certain functions belong to certain character categories. In Russian fairy tales there were barely seven character categories or roles (hero, helper, adversary, etc.)

When we develop further Propp's ideas, we arrive at the psychological significance of characters' functions (Péley, 2002; László, 2009). One of the major features of life stories is that their characters not only promote the plot by their actions, but also exert impact on the internal states of the person who narrates the story. These psychological functions as they appear in the life story are informative regarding personality development and personality states of the person. For instance, plot functions of helping or protecting can be interpreted in terms of psychological functions of defense and security, and it makes a difference whether this function is performed by the parents toward the child or the other way around, by the child toward the parents.

Characters' psychological functions and their linguistic operationalization

In our previous research (Péley, 2002; Péley and László, 2009) the following roles were identified in life stories: self, father, mother (parents), nuclear family, relatives, non-relatives. Identification of functions were based on characters' activity and attributes in story-situations. The following functions were identified:

- anti-model*: The character possesses features or commits deeds that becomes a negative example for the narrator and the narrator explicitly refers to these
- traitor*: according to the narrator, the character informed someone about an intimate thing, a 'common secret' without the awareness and approval of the narrator
- drug-friend*: relation to the character is connected to a drug
- leaver*: the character left the narrator because of divorce, removal, breaking-up.
- enemy*: according to the narrator, the character is against him/her, quarreled with him/her, or antecedents made the narrator dislike the character
- lost*: the character died
- adult associate*: the character reinforces the narrator in his/her adult identity, this function being more concrete than general supporting function
- threatening person*: the character does or says something that risks the narrator's psychical and/or physical existence
- restricting person*: the character inhibits the narrator in his/her physical or mental achievements
- model*: the character possesses features or deeds that appear as a model for the narrator
- non-caring*: according to the narrator, the character does not care for someone, does not assure the requirements of security, of existence
- non-supportive*: the character does not help the narrator achieve his/her aims, does not support him/her in difficulties, but at the same time does not inhibit him/her either, is not against him/her
- partner*: the character appears as a lover, or in intimate relationship
- helper*: according to the narrator, the character supports the narrator to achieve physical or mental aims, is an active participant and is attentive
- fellow*: the character and the narrator suffer the same thing
- anguishing person*: the character appears for the narrator as someone threatening. This can be caused by some concrete deeds and /or by a 'state' of the character
- supportive*: the character is actively present, paves the way for someone instead of supporting him/her, but does not helps him/her directly
- associate*: the character and the narrator do something together. It is not a passive way of being together, and some common (symmetric) experience appears
- competitor*: the character appears for the narrator as a rival: their aims are the same and they inhibit each other
- protector*: according to the narrator, the character provides security, protects from danger
- protégé*: the character requires the narrator's protection, and has the function of being someone who is protected by the narrator

These functions can be divided into two groups, having either negative versus positive features. Positive functions are: model, helper, supportive, and protector; negative functions are: anti-model, traitor, leaver, enemy, lost, threatening person, non-caring, and anguishing.

Linguistic operationalization of the functions has been performed in the LINTAG program (see Ehmann et al. 2007; Hargitai et al. 2007; Pohárnok et al. 2007; Pólya et al.

2007) It can be illustrated by the threatening function where the linguistic algorithm consists of four parts or phrases.

1. Verbs of physical insults or harm (torture, kick, hit, beat, hurt, etc.)
2. Certain prefixes in compound verbs, which express the lethal outcome of the action
3. Verbs of hostility directed toward the narrator by a character (quarrel, cry, threaten, humiliate, harass, etc.)
4. Approximation of any potentially threatening or harming object (e.g. gives a slash, lifts a gun, gets the wooden spoon, etc.)

At the time of the present study, the program has been able to identify character categories and character functions only separately, and matching the functions with character categories had to be performed manually. Nevertheless, a validity study examining whether character functions do reflect emotional regulation could be carried out. Results of this validity study may also provide support for the hypothesis about the relation between life narratives and self-development.

Sample and Method

Frame and Participants. As part of a more comprehensive project, correlational validity studies on character functions were based on a common textual corpus of various self-narratives. The total number of subjects involved in the study was 83 (29 men and 54 women; 50 young --18 to 35 years of age-- and 33 "old" --45 to 60 years of age--).

Textual corpus. The full corpus, collected from a total of 83 normal subjects on various autobiographical topics, was composed of 281,306 words (1,467,858 characters). The average length of individual texts was nearly equal, around 3500 (+/- 200) words per story.

Psychological tests. The autobiographic recalls were supplemented with a test battery for each subject. These involved, among others the *Hungarian version of the Beck Depression Questionnaire* (Kopp and Skrabski, 1990), and the Emotion Control and Impulse Control Factors of the *Hungarian version of the Big Five Questionnaire* (Rózsa et al., 1997).

Text processing. Lin-Tag software was used for computerized content analysis.

Statistical method. The subjects were distributed into low and high scorers (roughly along the lowest and highest quartiles) in the psychological questionnaires (BDQ, BFQ), and then two-sample t-tests were performed to explore the differences between the groups as to the frequencies of various psychological functions of the characters of their stories.

Results

Figure 1 shows that there are fewer positive functions in the stories of the high scorers in the Beck Depression Scale than in the stories of the low scorers [$t(39) = 2.07$ $p < 0.05$]. There was no difference in negative character functions.

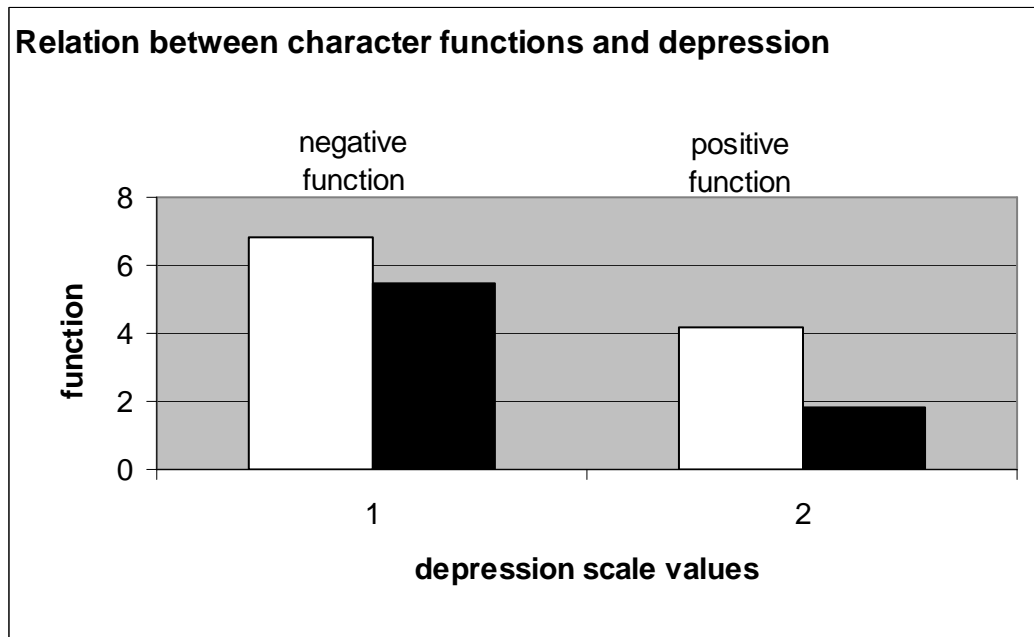


Figure 1. Relation between character functions and depression

For the Big Five Questionnaire we used the emotional control and impulse control scales. Here we also compared data of the low versus high scorers (lower versus upper quartiles) in each scale. Results are shown in Figure 2. Frequency of negative character functions were significantly higher for low scorers on both scales, i.e. subjects with low emotion control and low impulse control express more negative functions in their life stories than their counterparts do [$t(39) = 2.06$ $p < 0.05$ and $t(39) = 2, 087$ $p < 0.05$, respectively]. There were no gender or age differences in the sample.

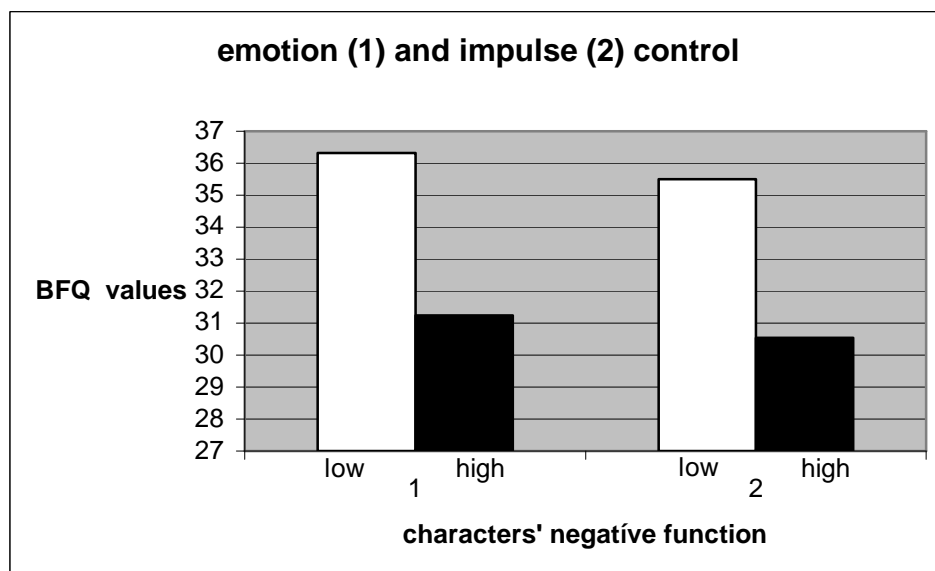


Figure 2. Relation between character function and emotion- and impulse control

Discussion

Although the linguistic algorithm for identifying character functions in life stories needs further refinements, results obtained with the current version seem to support the theoretical assumption that it measures self-states. Both dynamic and cognitive theories assume negative self-image and disturbed emotion regulation to be characteristic features of depression. Experiences of emotion regulation and self-experiences derive from the early mother-child interaction. In later stages of life, these experiences regulate how individuals perceive and interpret their interactions with other people and manage their interpersonal relations. When reconstructing their significant life episodes, the psychological functions they implicitly assign to their partners reflect this perception and interpretation and thereby their self-state. Thus, the low frequency of positive character functions in stories of subjects with high depression values shows hampered self-development. Self-developmental deviations are also indicated by the results which show that high frequency of negative character functions is related to low emotional- and impulse control.

References

- Bamberg, M. (2006). Stories: Big or small? Why do we care? *Narrative Inquiry*, 16 (1), 139-147.
- Barclay, C.R., & Smith, T. (1992). Autobiographical remembering: Creating personal culture. In M. A. Conway, D. C. Rubin, H. Spinnler, & W. Wagenaar (Eds.), *Theoretical perspectives on autobiographical memory* (pp. 1-28). Dordrecht, Netherlands: Kluwer Academic.
- Beebe, B. & Lachmann, F. M. (2002). *Infant research and adult treatment. Co-constructing interactions*. London: The Analytic Press.
- Bruner J., & Lucariello J. (1989). Monologue as narrative of the world. In K. Nelson (Ed.), *Narratives from the Crib* (pp. 73-97). Cambridge, Mass: Harvard University Press.
- Ehmann, B., Garami, V., Naszodi, M., Kis, B., & László, J. (2007). Subjective time experience: Identifying psychological correlates by narrative psychological content analysis. *Empirical Text and Culture Research*, 3, 14-25.
- Gergely GY., & Watson, J.S. (1996). The social biofeedback theory of parental affect-mirroring: The development of emotional self-awareness and self-control in infancy, *International Journal of Psycho-Analysis*, 77 (6), 1181-1212.
- Gergen, K. J. & Gergen, M. M. (1988). Narrative and the self as relationship. *Advances in Experimental Social Psychology*, 21, 17-56.
- Hargitai, R., Naszodi, M., Kis, B., Nagy, L., Bóna, A., & László, J. (2007). Linguistic markers of depressive dynamics in self-narratives: Negation and self-reference. *Empirical Text and Culture Research*, 3, 26-38.
- Kopp, M. & Skrabski, A. (1990). Az összehasonlító mentálhigiénés kutatás módszertana. (A methodology recommended for comparative mental health research.) *Végeken*, 2, 4-24. (In Hungarian)
- László, J., Ehmann, B., Pólya, T., & Péley, B. (2007). Narrative psychology as science. *Empirical Text and Culture Research*, 3, 1-13.
- Péley, B. (2002). Narrative psychological study of self and object representation with young deviant people. In J. László & W. Staiton Rogers (Eds.), *Narrative approaches in social psychology* (pp. 125-147). Budapest: New Mandate.

- Péley B., & László J. (2009). Narrative psychological study of self and object relation with young deviant people. *Empirical Text and Culture Research*, 4, 190-207.
- Pohárnok, M., Naszódi, M., Kis, B., Nagy, L., Bóna, A., & László J. (2007). Exploring the spatial organization of interpersonal relations by means of computational linguistic analysis. *Empirical Text and Culture Research*, 3, 39-49.
- Polkinghorn, D.E. (1995). Narrative configurations in qualitative analysis. *Qualitative Studies in Education*, 8, 5-23.
- Pólya, T., Kis, B., Naszódi, M., & László, J. (2007). Narrative perspective and the emotion regulation of a narrating person. *Empirical Text and Culture Research*, 3, 50-61.
- Propp, V. (1968). *The morphology of folk tales*. Austin: Texas University Press
- Ricoeur, P. (1984–1987). *Time and narrative* (Vol. I–III). Chicago: University of Chicago Press.
- Rózsa, S., Nagy, J., & Kö, N. (1997). *Rekonstruálható-e a Big Five magyarországi mintán?* (Is it possible to reconstruct the Big Five in Hungarian?). Budapest: ELTE University. (In Hungarian)
- Sarbin, T. R. (1986). The narrative as a root metaphor for psychology. In T. R. Sarbin (Ed.), *Narrative psychology. The storied nature of human conduct* (pp. 3-21). New York: Praeger.
- Siegel, D.J. (2001). Towards an interpersonal neurobiology of the developing mind: Attachment relationships, "mindsight" and neural integration. *Infant Mental Health Journal*, 22 (1-2), 67-94.
- Stern, D. N. (1995). *The motherhood constellation*. New York: Basic Books.
- Stern, D.N. (1989). Crib monologues from a psychoanalytic perspective. In K. Nelson (Ed.), *Narratives from the crib* (pp. 309-319.) London: Harvard University Press.
- Vincze O., Tóth J., & László J. (2007). Representations of the Austro-Hungarian Monarchy in the history books of the two nations. *Empirical Text and Culture Research*, 3, 62-71.
- Winnicott, D.W. (1990). On the contribution of direct child observation to psychoanalysis. In: *The maturational processes and the facilitating environment: Studies in the theory of emotional development* (pp. 109-114). London: Karnac Books.

Activity as a linguistic marker of agency: Measuring in-group versus out-group activity in Hungarian historical narratives¹

Katalin Szalai* and János László

*University of Pécs, Institute of Psychology
6. Ifjúság útja, Pécs, Hungary H-7624

Abstract

Agency is the ability to act effectively. It is a major component in social perception and identity studies. Life narratives and group narratives reflect agency of narrators in terms of semantic and grammatical features. It is assumed that implicit activity-passivity of verbs embodies important aspects of psychological agency. This study presents an automatic content analytic algorithm for mapping active and passive verbs. This algorithm is used to uncover Hungarian as well as out-group agency in a sample of historical narratives. Results are interpreted in terms of Hungarian national identity.

Agency

Agency has a wide range of psychological forms: e.g., capacity, expansion, power, dominance, autonomy, separation and independence. Harter (1978) defines the desire to control or have an effect on our environment as effectance motivation. Deci (1975) attributes inner control of actions to intrinsic motivation. DeCharms (1968) defines personal causation as a human disposition, implying intentional action for the sake of change. These phenomena all refer to the intention and desire to shape our physical and social environment.

Agency also plays a major role in Lazarus (1966) coping theory. Lazarus proposed a transactional model (in contrast with the earlier feature-model theory), in which both individuals and their environment are in a dynamic interaction and effect behavior in a stress-situation. Bandura's (1989, 1994) definition of self-efficacy, or personal efficacy, entails the idea that individuals are able to achieve their aims and control events happening in their life. The expectations that individuals have about their own efficacy are related to their coping: our belief in our own efficacy inspires us to invest more effort to achieve our aims and in a stress situation we feel less pressure or discomfort. Yamaguchi (2003) found correlations among control, personal efficacy and the psychological construct of autonomy: successful direct personal control leads to self-efficacy, which is the basis of the feeling of autonomy.

Vallacher and Wegner (1989) differentiate two poles of agency: low-level agents (How-People) who sees their actions in details, while high-level agents (Why People) focus on the meaning and the consequences of their actions. The high-level agent regards her or his

¹ Research was supported by OTKA T 49413 and NKFP6-00074/2005 grants to the second author

* szalai_katalin@freemail.hu

actions as the effect of mental factors such as motivation, rather than the effect of situational factors. Consequently, the actions of high-level agents appearing in different contexts indicate greater consistency, stability, and the self-organization than those of low-level agents. Actions of low-level agents are more impulsive, and tend to be described by using fewer mental states than those of high-level agents.

Not only individuals but also groups are seen as agents, as they are able to perform goal-directed behavior and have an effect on their environment. Hamilton (2007) distinguishes between two approaches to agency in group-perception research: one of them conceives agency as the capacity for efficient action (Abelson, Dasgupta, Park & Banaji 1998; and Brewer, Hong & Li, 2004), while the other one emphasizes the function of the mental states (Morris, Menon & Ames, 2001). The perception of a group's agency was measured by Spencer-Rogers, Hamilton and Sherman (2007) with four items: the group is 'able to influence others'; is 'able to achieve its goals', is 'able to act collectively' and is 'able to make things happen (produce outcomes)'. Kashima (2005) assessed perceptions of agency with nine items mapping mental states (beliefs, desires and intentions) which, according to him, are the basis of the group-agency.

Agency and identity

At least in Western cultures, agency is an important component of personal and social identity. In order to arrive at a well organized and adaptive adult identity, people have to acquire autonomy, which is reflected in their agency in life events (McAdams, 2001). Current narrative models of identity reconstruct personal identity from life stories (Bamberg and Andrews, 2004; Brockmeier and Carbaugh, 2001; Freeman, 1993; McAdams, 1985). Similar to individual identity, group identity can also be reconstructed from narratives about the group's past. Representations of history reflect psychological characteristics of national identity such as stability or vulnerability, strength or weakness, autonomy or dependency, etc. (László, 2008; Liu and László, 2007). Distribution of agency between in-groups and out-groups seems to be a sensitive indicator of the above identity states.

Linguistic markers of agency

Agency can be best expressed by voice. Active voice (e.g. "Hungarians occupied their land") carries higher level agency than passive voice (e.g. "The land was occupied by Hungarians"). In languages, where the passive voice is seldom or not at all used, general subjects instead of nouns or personal pronouns may express passivity.

Another means of communicating agency is thematization. It shows who is the focal character in the event presented by the text. The theme is presented first in sentences, it has the grammatical role of surface subject and therefore it frequently is associated with passive verbs. In "John greeted Mary" the theme is John, however in "Mary was greeted by John" the theme is Mary. Although languages may differ in how thematization is performed, thematic information universally occupies the initial position in sentences (Semin, 2000). Turnbull (1994) showed that the thematic character carried higher responsibility for moral events, therefore her agency is also perceived higher.

There is, however, a third means, which is less connected to grammatical structures than to implicit semantics of verbs (see e.g. Brown and Fish, 1983). Similar to implicit causality, verbs have an implicit activity versus passivity meaning. We utilized this aspect of

agency by constructing a computer algorithm, which maps active and passive verbs in texts. It is in accord with Abelson et al., (1998) and Brewer et al., (2004) who derive agency from effective actions.

The analytic device

The computer algorithm was developed using the NooJ linguistic development environment. NooJ is able to perform morpho-syntactical analysis (Silberztein, 2008). Five judges classified the dictionary of the ten thousand most often-used verbs into different verb categories. A verb was considered active when it referred to an intentional act, the action had an agent, and the action had an effect on the environment.

*“The Hungarian troops **occupied** Bácska, the Baranya triangle and Muraköz.”*

*“Rákóczi **turned to** the French king for help.”*

*“Hungary **took a hand** in redeeming the debts of the empire.”*

*“He **chose** suicide 3 April in 1941.”*

Verbs indicating change in state or action were classified as passive verbs. These actions refer to change outside the individual’s control, e.g., changes in physical circumstances, or transcendental changes. For instance:

*“Later, the centres of immigration **grew up** in Paris and London.”*

*“By 1917 revolts and escapes **occurred** in the military.”*

*“On account of these, our country **got** under totally new life conditions.”*

*“Thirty percentages of the Hungarian habitants, so 3.3 million Hungarian people, **became** under control of other states.”*

For isolating languages, such as English, it suffices to compile a dictionary. However, morphological complexity prevents using simple dictionaries. In agglutinative languages, such as Hungarian, different word forms cannot be handled by lexicons. For this reason not only words but also grammatical relations between them should be analyzed. NooJ has the capacity for building local grammars around words. In the following figure, the graph of the activity-passivity local grammars is shown. Squares contain further graphs.

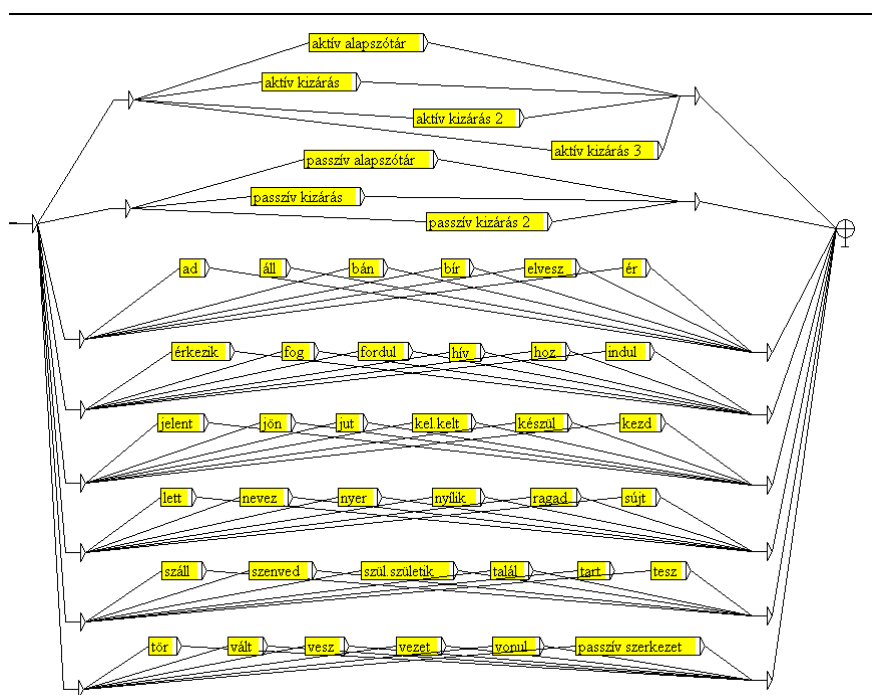


Figure 1: The main graph of activity-passivity module

Local grammars are able to find and identify the different cases and inflections of the verbs with appropriate output. The graph of activity-passivity, consists of local grammars of 941 active and 230 passive verbs.

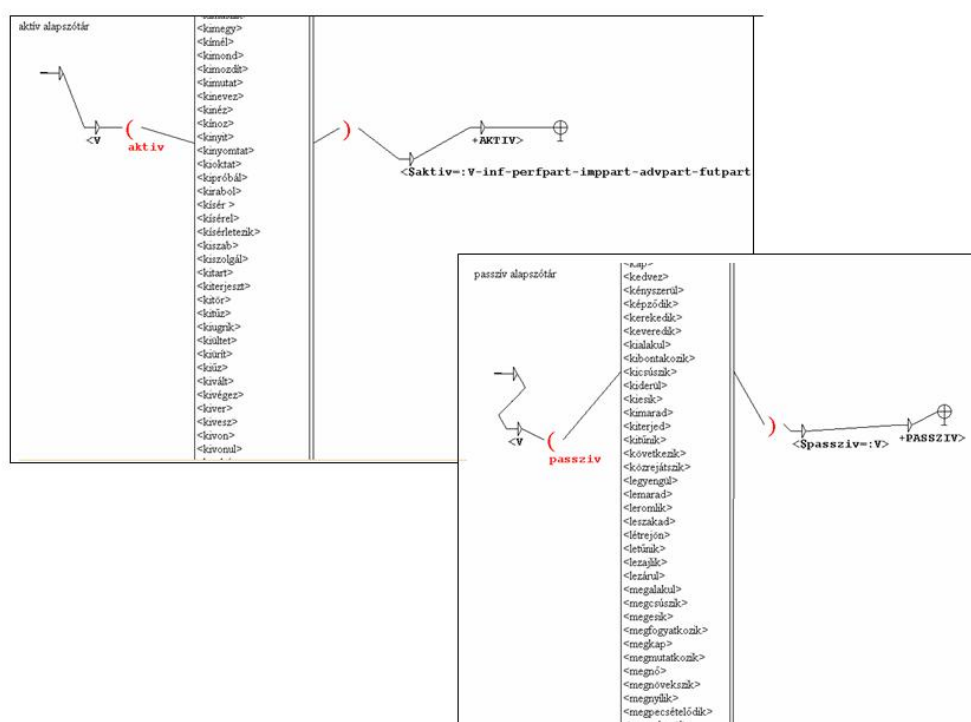


Figure 2: An item of the active and passive dictionary

Within the main graph, there are local grammars whose constructions are different from the ones in the basic dictionary. Their functioning is demonstrated by the Hungarian verb ‘fordul’ (to turn) on Figure 3. It is shown how the algorithm divides the active and passive case of the verbs with the help of the context and the government of the verbs. Figure 4 shows a sample list of hits in the central column, together with their concordances in the left and right columns.

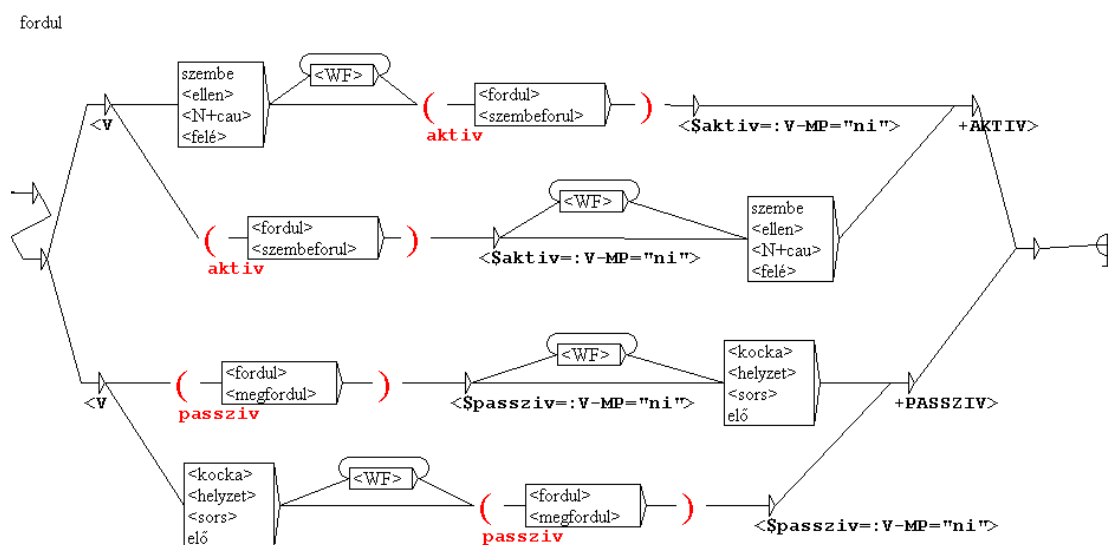


Figure 3: The graph of the Hungarian verb 'fordul' (to turn)

Text	Before	Seq.	After
szerződés aláírása után „örök barátja”	ellen fordul/<V+AKTIV>		. A dilemma megoldására egyetlen lehetőséget
az 1937-es határok mögé és	szembe kell fordulni/<V+AKTIV>		a németekkel.) Közben Horthy hozzálátott
tavaszán a német terjeszkedés Skandinávia	felé fordult/<V+AKTIV>		. Norvégia gazdasági jelentőségét az adta

Query 3/3

Text	Before	Seq.	After
Puchner Antal császári tábornok ekkor	fordult először segítségért/<V+AKTIV>		a Havasalföldön állomásozó orosz seregekhez
adó Alföldre kitorjón. A hadi	helyzet alapvetően megfordult/<V+PASSZIV>		, amikor erősítésekkel megérkezett a lengyel

Query 2/2

Figure 4: The concordance tables of the hits in NooJ program

The algorithm is able to detect linguistic structures of activity-passivity in text, and enables quantitative analysis of each character's activity. The higher is the activity-passivity ratio (i.e., the more active idioms are used relative to the passive ones), the more the character is presented as an efficient actor in the narrated events and the more she is having effect on her environment. Conversely, the more passive expressions the narrative uses relative to the passive ones for a character, the more it emphasizes her passivity and incapacity in action.

Distribution of activity-passivity in Hungarian historical narratives

In our study we used two text corpora: 1. Textbooks used in primary and secondary schools (approximately 150 000 words). 2. Folk narratives collected from a sample of 500 subjects, stratified by age, education and ethnic background (approximately 64 thousand words). Subjects were asked to tell narratives, which they found the most positive and the most negative events of the Hungarian history. Events in the two historical text corpora were matched with each other.

Table 1 shows the examined historical events.

Table 1: The examined historical events

The examined historical events		
Positive	Negative	Positive and negative
Conquest of homeland	Mongolian invasion	Turkish rule
Establishment the state	Paris peace treaty (Trianon)	Habsburg rule
1989 system's change	World War two	Hungarian revolution of 1956
	Holocaust	

The activity-passivity algorithm was run on both text corpora in order to see to what extent Hungarians (in-group) are depicted as active agents and passive sufferers in the historical narratives and what are the events in which Hungarians turn up active as opposed to the

events where they are depicted as passive. Similar questions were raised regarding the non-Hungarian participants.

Results

Results are shown in Figure 5 in terms of the rate of active and passive frequencies. Data are presented separately for positive and negative historical events. Higher values indicate that more active verbs were used in the description of the events. Moreover, the more the index approaches 1, the more passive verbs occurred in the given historical story. If the index is less than 1, it means that they used more passive expressions than active ones.

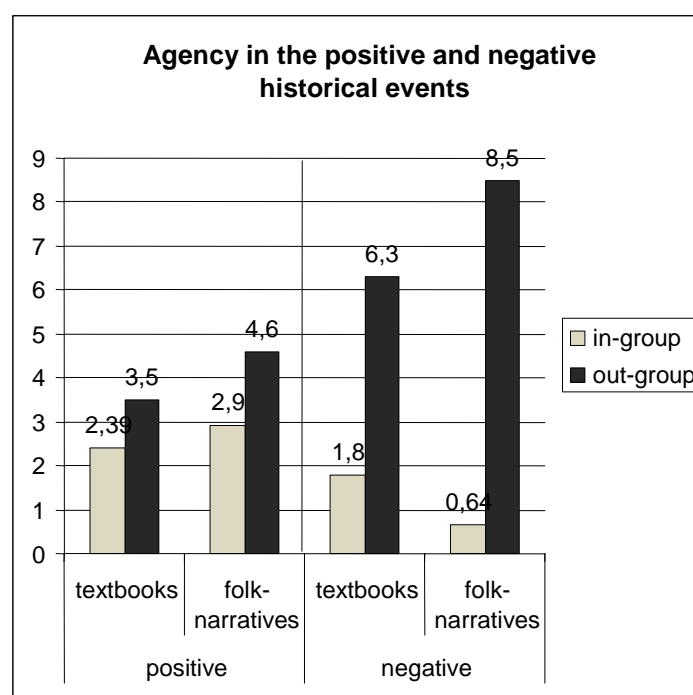


Figure 5: The average of the agency-indices divided them into positive and negative historical events

Overall, the results show that the out-group appears to be more active than the in-group for both positive and negative events. In the positive events, Hungarians show more activity than in the negative ones, and conversely, the out-group comes up as more active in the negative events. These differences are more pronounced in folk-narratives than in historical textbooks. Results are in accord with the observation of social psychological attribution research: people tend to ascribe their success to internal factors, whereas success of others is attributed to external factors (Jones and Nisbett, 1972). Still, the difference in our results, particularly in folk narratives, is remarkable.

In Figure 6 and Figure 7 the indices found in the different historical events are shown. In the history textbooks, the index of one single event (Paris peace treaty- Trianon) is lower than 1. In the case of folk-narratives, indices of five events (Mongol invasion, Turkish rule, Paris peace treaty- Trianon, World War II., and repression of the '56 revolution) are below 1. Note that all these events are negative, i.e. they are related to defeats. Even in the case of

positive events, in-group agency is relatively low except events in the distant past (conquest of land, establishing the state). In the most recent positive event, system's change, the agency index barely exceeds 1.

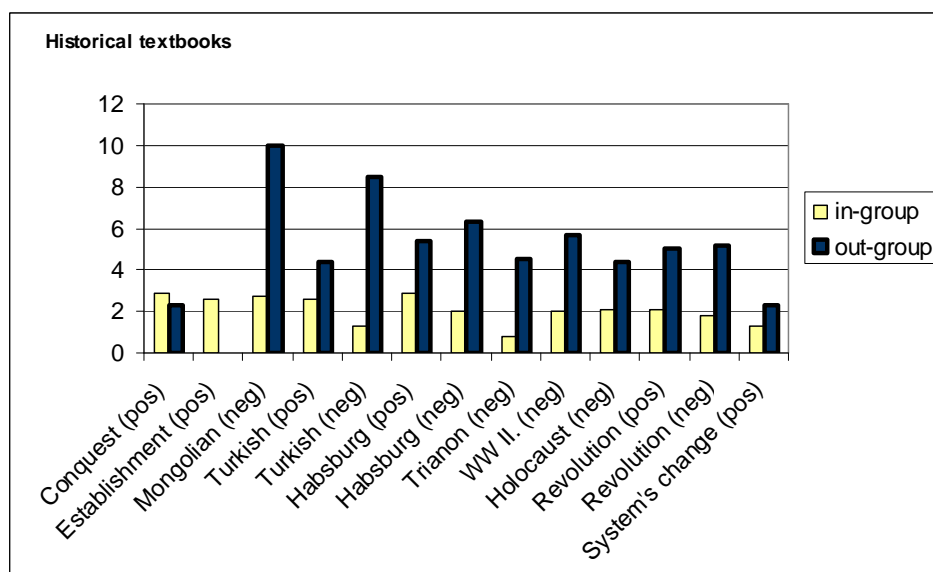


Figure 6: Agency in historical textbooks separated into in-group and out-group

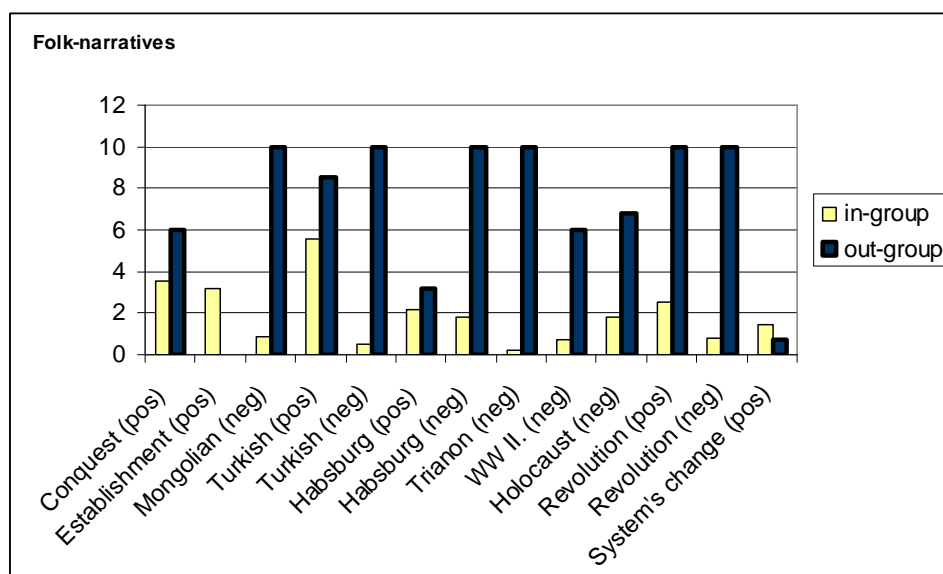


Figure 7: Agency in folk-narratives separated into in-group and out-group

Discussion

Several historical accounts argue that the Paris peace treaty (Trianon treaty) had been a traumatic experience for the Hungarian national identity after the WW 1 and this trauma persists even today (Romsics 2001, Ormos 1983). The fact that both history text books and folk narratives depict Hungarians in this event as helpless and defenseless underpins this interpretation.

Whereas history text books endow Hungarians with relatively more agency, in folk narratives representing more vivid forms of collective memory Hungarians turn out to be at the mercy of the invasions of bigger empires.

Defenselessness raises the question of responsibility or rather the question of irresponsibility, as the experience that we cannot have effect on our fate, can easily go together with the feeling that we are not responsible for our life. Historians claim that several generations of Hungarians, have grown up with this feeling. They have not got used to the independency and the frames of democracy (Ormos, 2005). Our results support this observation. In the case of the system's change, we found a high degree of passive verbs. Although the agency-index is higher than one (more active expressions than passive ones occur in the texts), the agency-index is the lowest among the positive events both in the folk-narratives as well as in the stories of the textbooks. This result shows that this important event, which could have served as strengthening Hungarian national identity, is represented as an event out of our own efficacy or our own activity.

Summing up, we can say that the stories about the Hungarian history -both the professional ones, which are the parts of history textbooks, and the stories of the naïve historians, - present identification patterns in which Hungarians -apart from events in the distant past-- appear with reduced capacity of action as compared to other nations. Low level of agency may weaken their responsibility for their fate and the realistic appraisal of their current situation. Contemporary history text books seem to alleviate this identity construction, but their impact on the common forms of collective memory is still in question.

References

- Abelson, R.P., Dasgupta, N., Park, J., & Banaji, M.R. (1998). Perception of the collective other. *Personality and Social Psychology Review*, 2, 243-250.
- Bamberg, M., and Andrews, M. (2004). Introduction. In M. Bamberg & M. Andrews (Eds.), *Considering Cuonter-Narratives: Narrating, resisting, making sense*. Amsterdam: John Benjamins.
- Bandura, A. (1989). Perceived self-efficacy in the exercise of personal agency. *The Psychologist: Bulletin of the British Psychological Society*, 2, 411-424.
- Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (Vol. 4, pp. 71-81). New York: Academic Press.
- Brewer, M.B., Hong, Y., & Li, Q. (2004). Dynamic entitativity: Perceiving groups as actors. In V. Yzerbyt, C. Judd & O. Corneille (Eds.), *The psychology of group perception: Perceived variability, entitativity, and essentialism* (pp. 25-38). New York: Psychology Press.
- Brockmeier, J., & Carbaugh, D. (2001). *Narrative and identity: Studies in autobiography, self and culture*. Amsterdam/Philadelphia: John Benjamins.
- Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, 14, 237-273.
- deCharms, R. (1968). *Personal causation*. New York: Academic Press.
- Deci, E.L. (1975). *Intrinsic motivation*. New York: Plenum.
- Freeman, M. (1993). *Rewriting the self: History, memory, narrative*. London: Routledge.
- Hamilton, D.L. (2007). Understanding the complexities of group perception: Broadening the domain. *European Journal of Social Psychology*, 37, 1077-1101.
- Harter, S. (1978). Effectance motivation reconsidered: Toward a developmental model. *Human Development*, 6, 34-64.

- Jones, E.E. & Nisbett, R.E. (1971). The actor and the observer: Divergent perceptions of the causes of behavior. In E.E. Jones (Ed.), *Attribution: Perceiving the Causes of Behavior* (pp. 79–94). Hillsdale, NJ: Lawrence Erlbaum.
- Kashima, H.Y. et al. (2005). Culture, essentialism, and agency: Are individuals universally believed to be more real entities than groups? *European Journal of Social Psychology*, 35, 147–169.
- László, J. (2008). *The Science of stories: An introduction to narrative psychology*. New York and London: Routledge.
- Lazarus, R.S. (1966). *Psychological stress and the coping process*. New York: McGraw-Hill.
- Liu, J.H., & László, J. (2007). A narrative theory of history and identity.: Social identity, social representations, society and the individual. In G. Moloney, & I. Walker (Eds.), *Social representations and history* (pp. 85–107). Basingstoke: Palgrave-Macmillan.
- McAdams, D.P. (1985). *Power, intimacy, and the life story: Personological inquiries into identity*. New York: Guilford Press.
- McAdams, D.P. (2001). *Coding autobiographical episodes for themes of agency and communion*. URL: http://www.sesp.northwestern.edu/docs/Agency_Communion01.pdf
- Morris, M.W., Menon, T., & Ames, D.R. (2001). Culturally conferred conceptions of agency: A key to social perception of persons, groups, and other actors. *Personality and Social Psychology Review*, 5, 169–182.
- Ormos, M. (1983). *Padovától Trianonig, 1918–1920*. (From Padova to Trianon, 1918–1920). Budapest: Kossuth Könyvkiadó.
- Ormos, M. (2005). *A történelem vonatán. Európa és Magyarország a 20. században*. (On history train. Europe and Hungary in the 20th century). Budapest: Múlt és Jövő Kiadó.
- Romsics, I. (2001). *A trianoni békeszerződés*. (The Trianon peace treaty). Budapest: Osiris Kiadó.
- Semin, G.R. (2000). Language as a cognitive and behavioral structuring resource: Question-answer exchanges. *European Review of Social Psychology*, 11, 75–104.
- Spencer-Rogers, J., Hamilton, D.L., & Sherman, S.J. (2007). The central role of entitativity in stereotypes of social categories and task groups. *Journal of Personality and Social Psychology*, 92, 369–388.
- Turnbull, W. (1994). Thematic structure of descriptions of violent events influences perceptions of responsibility: A thematic structure effect. *Journal of Language and Social Psychology*, 13 (2), 132–157.
- Vallacher, R.R. & Wegner, D.M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology*, 57 (4), 660–67.
- Yamaguchi, S (2003). A kultúra és a kontrollal kapcsolatos orientációk. In L.L.A. Nguyen & M. Fülöp (Eds.), *Kultúra és pszichológia*. (Culture and Psychology) (pp. 386–408). Budapest: Osiris Kiadó.

In-group versus out-group intentionality as indicators of national identity¹

Réka Ferenczhalmy* and János László**

University of Pécs, Institute of Psychology
6. Ifjúság útja, Pécs, Hungary H-7624

Abstract

Intentionality plays a significant role in the evaluation of actions. It is related to agency, to attribution of responsibility and of control. The study presents an automatic content-analytic algorithm for identifying linguistic markers of intention, constraint and possibility in the text. This algorithm was used to uncover both the Hungarian and the out-group intentions in a sample of Hungarian historical narratives taken from historical text-books and from a pool of folk-narratives. Results are interpreted in terms of the Hungarian national identity.

Keywords: intention, agency, historical narratives, national identity

Intention

The present research does not cover the complete, broad concept of intentionality, but examines it in a narrower sense, referring to purpose. For clarification of the meaning of purpose we first need to differentiate events happening on purpose and not on purpose. Considering it from the phenomenological perspective, this spectrum ranges from the absolutely unwitting reflexes, through the subintentional (e.g., swinging our legs during a conversation) and preintentional movements (e.g., walking), to the purposive behaviour.

A motion or behaviour has to be expedient and purposive to be considered become an activity. Thus, there are no unpremeditated actions (nevertheless, we can talk about unintentional motion and not aforethought consequences of action). To define what makes an action purposive, that is, what leads to its being an action, and what precisely we are aware of and to what extent during the purposive action, we need to consider – besides the concepts of intention and action – two aspects of agency.

The first one, the experiential feeling of agency, appears on the level of pre-reflective mind together with the action (when I feel that I am moving, though I do not know my motion in every detail precisely). The second aspect is the attribution of agency, which is based upon the first one, such that I attribute the action to myself and can even give account of it.

¹ Research was supported by OTKA T 49413 and NKFP6-00074/2005 grants to the second author. Correspondence should be addressed to the first author.

* ferenczhalmy@gmail.com

** laszlo@mtapi.hu

Searle's concept of 'intention-in-action' can be connected to the notion of agency. This relates to actions accompanied by explicit decisions (e.g., if someone is determined to go round to the neighbours', then this intention will be reflected in standing up, opening the door, etc., even if the pragmatic purpose was to step over to the neighbours'), (Gallagher & Zahavi, 2008).

In the present research we study intentionality as the attribution of purpose, which specifically postulates an explicit decision and is bonded to reflective consciousness.

Narrative, history and group identity

Regarding our understanding or thinking about the world, Bruner (1986/2005, 2001) differentiated two basic modes: pragmatic (logical-scientific) and narrative thinking. The former rests on the application of logic operations and its content must meet criteria of verification. The latter stands for thinking in narratives, where life-likeness and probability are crucial for verification. Thus, the understanding of others' (or one's own) actions, feelings, intentions, and of the interrelation of these, is constructed in term of narrations.

Information is coded on two distinct levels of narration: on a descriptive level, containing the bare description of the occurrences; and on a psychological level, involving the perspective, relation and evaluation of the narrator. At this level, psychological aspects appear: desires, intentions, emotions and so on, implying some kind of interpretation of the events. On this psychological level the role of interpretation and construction has to be emphasized, as the same situation or event might be reported variously by different individuals, and also one person may reinterpret the given situation at any time.

Identity is understood as a complex of continuously recombined (life)events, in which based on the events of the past (descent, roots, life experiences) and the present, possible future may be projected. The narrative approach to identity appears in several theories (Ricoeur, 1992; McAdams, 1985, 2006; Gergen & Gergen, 1988).

Personal identity means stability and continuity of the self in the midst of constant changes. (e.g., Erikson, 1968). However, one can also talk about social identity, which, according to the approach of Tajfel (1981, Tajfel and Turner, 1986) refers to the uniformity of the individual with others, with a group. Personal identification with the group, accepting its norms and values and becoming member of this group, provides self-reinforcement and safety for the individual.

The present research considers social identity by accessing history books as determining social representations in the formation of social identity. These representations, according to the theory of Moscovici (1961), emerge through a constructive process and determine social reality (on cultural, social and individual levels); in other words, determine the contents of everyday discourses and their meanings. Information therefore extends beyond the objective knowledge, involving the attribution of meaning, assigning the possible evaluation and relations, in close connection with the actual group processes, interests and goals. The social representations are consequently rather important in respect of social identity. (László & Liu, 2007; László, 2008)

Scientific narrative psychological approach

The narrative psychological approach (László, 2008) attempts to capture the characteristics of experience organization appearing in narratives, and at identifying the linguistic features - and their patterns of appearance - relevant for meaning construction.

In this research, historical texts, being significant social representations regarding national identity, are examined from a narrative psychological perspective. This study considers the features pertaining to intentionality that appear on the interpretative level within the narration of certain events, the results of which will be explicable in terms of national identity.

Intention, constraint and possibility in the text

Those actors (or group of actors), to whom during the narration of a story we assign numerous intentions, are perceived as capable of acting in order to reach given goals, in contrast to those who are attributed few intentions but many of constraints. By examining intention, constraint and possibility, the context of action becomes comprehensible, possibly playing a role in the evaluation and attribution of such dimensions as control, responsibility and efficiency, which are in turn significant in the evaluation of the acts and story of an individual or a group. Similarly, high intention level and low constraint level in self-narratives or group narratives indicate stable, well-organized and autonomous identity.

Intention in the text

The appearance of intentionality, that is of the attribution of purpose might occur on several levels in the text. In some respect, all action verbs can be considered intentional (recall the concept of intention-in-action above), but in our research we only identify those cases, where - in a phenomenological sense - explicit intentions are present, which - according to Brunner's distinction - appear on the psychological level of the narrative.

Methodology and means of analysis

Our research team has been developing dictionary-based local grammars for narrative psychological content-analysis within the NOOJ integrated language-analysis environment. This software enables morpho-syntactic, that is textual level analysis. Beyond the elaboration of the content relevant dictionaries for the examination, the identification of grammatical structures is enabled also with the help of local grammars, both for general (e.g., verb + infinitive) and for concrete cases (e.g., any form of the verb to go + infinitive). Consequently, not only concrete words can be searched for, but also word forms and various text environments, which are important regarding meaning assignment. The following graphs and examples help to explain this.

The complete local grammar of intention consists of the following units:

1. Local grammar of intentional adverbs, adjectives and postpositions

The graph consist of two dictionaries, one of the adverbs of manner and one of the adjectives. Those adverbs of manner belong here, which carry intention linked to whichever active operative verb (e.g., someone acts voluntarily or involuntarily, willfully or accidentally). Adjectives might carry the same information by being linked to otherwise intentionally neutral nouns (e.g., whether the action was intended, planned or random).

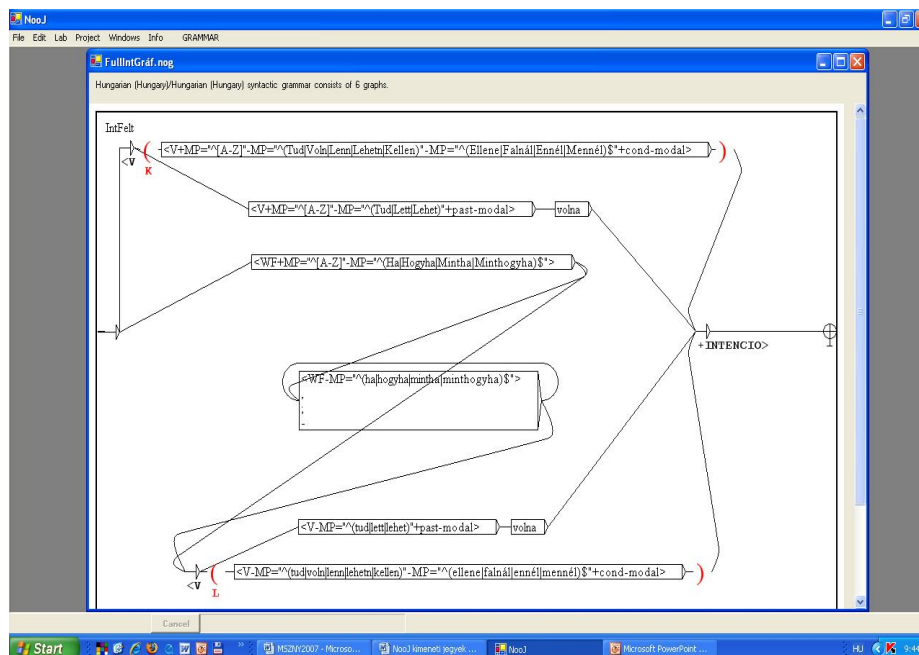


Figure 2. Local grammar of conditional mood

5. Local grammar of final subordinate structure of sentence

This graph serves the identification of intentional cases where the goals of concrete action are indicated in the text (e.g., one does something to be happy).

Local grammar of constraint

This local grammar identifies forms of constraint that are expressed linguistically.

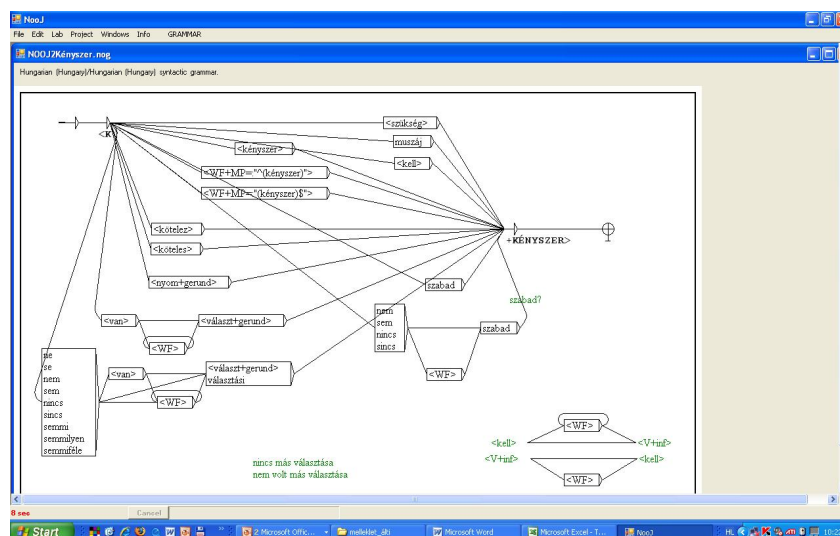


Figure 3. Local grammar of constraint

through the analysis of these texts information might be obtained about national identity.

Assumptions

- If a group is attributed more intention and less constraint in a text, this indicates stable and autonomous identity.
- In the narration of traumatic events, higher levels of possibility and intention attributed to the in-group can be seen as indicator of elaboration. It also implies stable identity.

Questions raised in the study

- A. 1. In the narration of historical events, what level of intention characterizes the Hungarian group and to what extent does constrain appear?
2. In the same narrations, what measure of intention and constraint is characteristic for the other group (out-group)?
- B. 1. How general are the measures of intention and constraint attributed to the Hungarians considering the descriptions of various events?
2. How general are the measures of intention and constraint attributed to the other group considering the descriptions of various events?
- C. 1. How general is the measure of possibility attributed to the Hungarians considering the descriptions of various events?
2. How general is the measure of possibility attributed to the other group considering the descriptions of various events?

Results

Relative frequencies of each content category were calculated, hence making results concerning different events comparable.

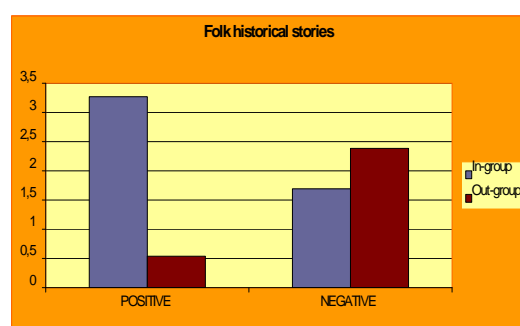


Figure 5. Intentionality of the Hungarian in-group and various out-groups in positive and negative historical events in the folk narrative corpus

Results concerning intentionality of the Hungarian in-group and various out-groups in positive and negative historical events in the folk narrative corpus are shown in Figure 5.

When describing positive events the intention of the in-group is high, but that of the out-group is remarkably low. For the negative events, the intention level of the out-groups is high, but that of the in-group is relatively low.

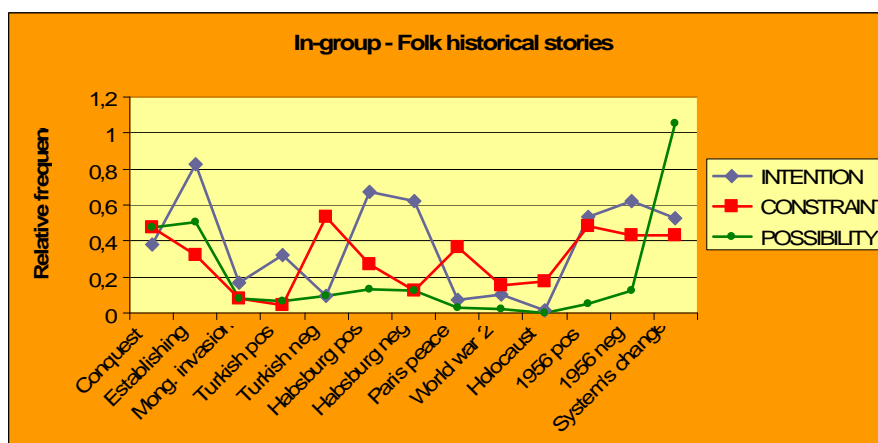


Figure 6. Intentionality, constraint and possibility of the Hungarian in-group in the folk narrative corpus

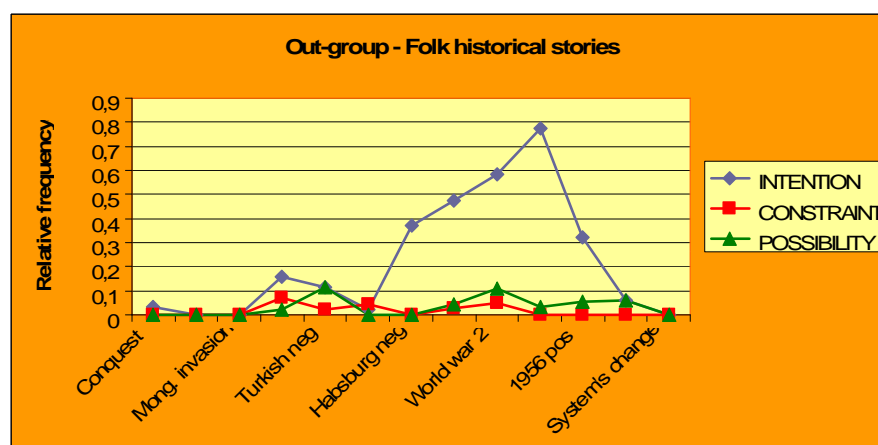


Figure 7. Intentionality, constraint and possibility of the out-groups in the folk narrative corpus

Considering the three factors of intentionality, constraint and possibility of the in-group and out-groups in the folk narrative corpus (shown in Figure 6 and 7), the in-group is distinct that such for positive events the high level of intention is combined with a lower level of constraint and generally higher level of possibilities, while for negative events the ratio of intentions and constraints shows the opposite pattern, whilst the level of possibilities decreases.

In the case of the out-group, the level of intentionality is remarkably low for positive events, and high for negative ones. Constraint and possibility are both low (see Figure 7).

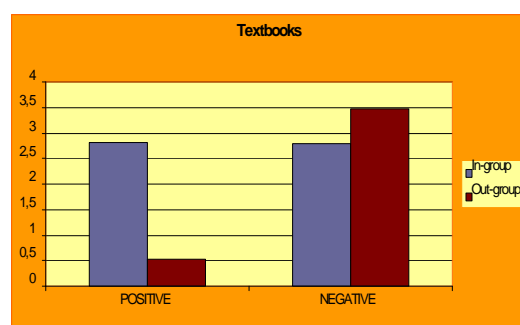


Figure 8. Intentionality of the Hungarian in-group and various out-groups in positive and negative historical events in the history book corpus

In the history textbooks, the intention attributed to the in-group is relatively high for both positive and negative events (see Figure 8.). The intention of the out-groups is remarkably low when describing positive and high when describing negative events.

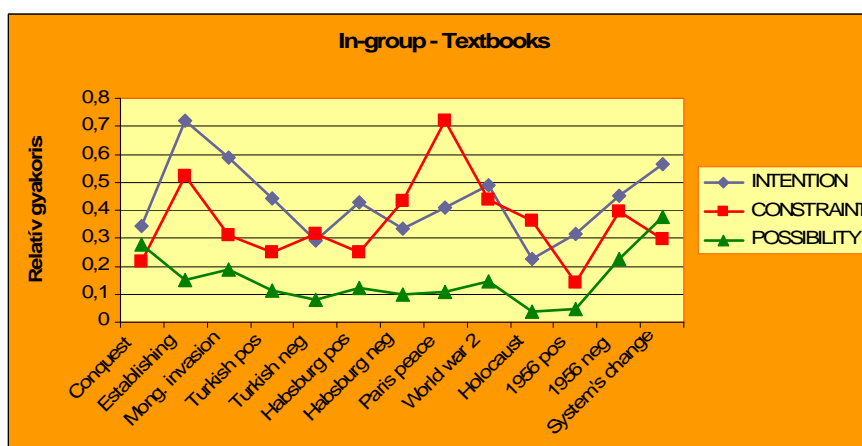


Figure 9. Intentionality, constraint and possibility of the Hungarian in-group in the history book corpus

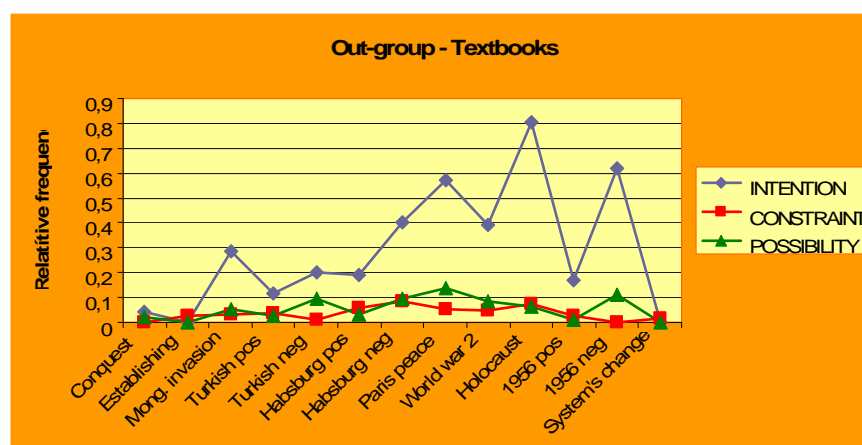


Figure 10. Intentionality, constraint and possibility of the out-groups in the history book corpus

The history book results show the same pattern as in the folk historical corpus. For the in-group, in the case of positive events the level of intention is combined with a lower level of constraint and generally higher level of possibilities. In the case of negative events, the ratio of intentions and constraints reverses, whilst the level of possibilities decreases. The level of possibility increases only in the narrations referring to events of the recent past. (see Figure 9.)

In the case of the out-group, the level of intentionality is remarkably low for positive events, and high for negative ones. In neither case do the results of possibilities and constraints change in the various events' narratives. (see Figure 10.)

Discussion

To sum up our results, we can conclude that a relatively high level of intention is attributed to the in-group for positive historical events and a relatively low level for negative events in both corpora. High level of intention generally occurs with low level of constraint and high level of possibilities. When the level of intention is low, high level of constraint and low level of possibilities occur.

Considering the out-group, the pattern of the attribution of intention is the opposite. High level of intention is attributed in the case of the positive historical events, and low level in the case of the negative events. This tendency appears in both corpora. There is no remarkable pattern in the occurrence of intention constraints and possibilities.

What inferences can be made regarding Hungarian national identity on the basis of these results? Although it is a well-known self serving bias (see e.g., Forgas, 1985) in attribution research that success or victory is attributed to internal, whereas failure or defeat to external causes, the degree of this bias appearing in the presented folk-narratives is remarkable. In negative historical events Hungarians are depicted as not having intentions, not contemplating possibilities, and facing only constraints. In these events outgroups are seen as highly intentional and having no constraints. This pattern of results is the most conspicuous with the Paris peace treaty subsequent to WW1, which seems to be the most traumatic event in Hungarian history in the past century. In accord with our assumptions, extremely low level of intentionality and high level of constraint in negative events suggest that both history textbooks and folk narratives reflect on vulnerability and instability of the Hungarian national identity. In the same way, the low level of in-group possibilities and intentions as opposed to the extremely high level of out-group intentions in the most traumatic event of the 20th century suggest that meaning of this trauma has not progressed too far. It also indicates instabilities in Hungarian national identity. There are, however, two positive features for the Hungarian national identity in our results. First, the above tendencies are expressed to a lesser extent in school books than in folk narratives. This suggests that elaboration of the national traumas has at least begun on this level of historiography. Second, in-group intentionality increases in the events of the recent past whether negative ('56 revolution) or positive (system's change). These results point again to a stabilization of national identity.

References

- Bruner, J. (1986). *Actual mind, possible worlds*. Cambridge, MA: Harvard University Press.
- Bruner, J., & Lucariello, J. (1989). Monologue as narrative recreation of the world. In *Narratives from the crib* (pp. 73-97). Cambridge, MA, US: Harvard University Press.
- Forgas, J. P. (1985). *Interpersonal behaviour. The psychology of social interaction*. Oxford: Pergamon.
- Gallagher, S., & Zahavi, D. (2008). *The phenomenological mind. An introduction to philosophy of mind and cognitive science*. New York: Routledge.
- Gergen, K. J., & Gergen, M. M. (1988). Narrative and the Self as Relationship. *Advances in Experimental Social Psychology*, 21, 17-56.
- László, J. (2008). *The science of stories: An introduction in narrative psychology*. New York and London: Routledge.
- Liu, J. H., & László, J. (2007). A narrative theory of history and identity: Social identity, social representations, society and the individual. In G. Moloney & I. Walker (Eds.), *Social representations and history* (pp. 85-107). Basingstoke: Palgrave-Macmillan.

- McAdams, D. P. (1985). *Power, intimacy, and the life story: Personological inquiries into identity*. New York: Guilford Press.
- MacAdams, D. P. (2006). *The redemptive self: Stories Americans live by*. New York: Oxford University Press.
- Ricoeur, P. (1992). *Oneself as another*. Chicago: University of Chicago Press.
- Tajfel, H. (1981). *Human groups and social categories*. Cambridge: Cambridge University Press.
- Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. G. Austin (Eds.), *The psychology of intergroup relations* (pp. 7-24). Chicago: Nelson-Hall.

Personality and the Familial Unconscious in Szondi's Fate-Analysis

Enikő Gyöngyösiné Kiss¹

University of Pécs, Institute of Psychology, Department of Personality, Developmental and
Clinical Psychology, 6. Ifjúság útja, Pécs, Hungary H-7624

Abstract

Szondi considered his fate-analysis as belonging to depth psychology and he thought that it would be a contact between Freud and Jung. Szondi supposed that between the personal and collective unconscious there is a third, the familial unconscious. The familial unconscious is originally based on genealogy. The descendent gets his genetic structure from his ancestors which determines the possibilities of his life and fate. Consequently, the goal of fate-analysis is the scientific investigation of human destiny. Szondi's phrase was that „*Fate is always on the move*”. According to Szondi, fate has not only a compulsive part but also a freely chosen part, and he distinguished the components of compulsive and selective fate. The selective fate is directed by the ego (Pontifex-ego) and the spirit. The aim of fate-analytical therapy is to find a much happier fate for the patient as he lived the compulsive fate earlier. The readers of the journal could read about the Szondi test in the previous number by Rita Hargitai. This article presents a short sketch of Szondi's fate-analytical theory and practice.

Keywords: Familial unconscious, compulsive fate, selective fate, choices of partner-, friend- profession-, illness- and form of death, stand-taking ego (Pontifex oppositorum, Pontifex ego), fate-analytical therapy.

Introduction

Lipót (Leopold) Szondi (1893-1986) was a psychiatrist of Hungarian origin, who later in his life settled and worked in Zurich (Bürge-Meyer, 2000; Bürge-Meyer, Gyöngyösiné Kiss 1994). His fate analysis is considered among the depth analytical schools of the 20th century. As Szondi defined it, fate analysis is 'genetics introduced to psychoanalysis'. Szondi was familiar with Freud's psychoanalytic writings, and besides this influence, his views were based on the findings of contemporary international twin-research, neuro-endocrinology and the complex examination of mentally handicapped children (the impact of biological and environmental factors considered together). He created a human drive theory, in which individuals' present and future behavior, their characters and the existential possibilities of their fate are all determined by the dialectics of the opposing drive-pairs of four drives. He also compiled the so-called Szondi-test, which enables researchers to assess the individual differences in manifest and latent drive needs; to describe and diagnose the functioning of the

¹ E-mail address: gyongy@btk.pte.hu

healthy and the pathological personality (Szondi, 1947, 1952; Szondi, Moser, Webb, 1959; Déri, 1949).

To complement the Freudian concept of the individual unconscious, and Jung's idea of the collective unconscious, Szondi introduces the notion of familial unconscious to describe the role of individuals' familial ancestry grounding their choices. In Szondi's conception, people's fate is shaped by their choices. The choices of partner-, friend-, profession-, illness- and form of death are pivotal among these (Szondi, 1944).

The ancient figures of the familial unconscious refer to the determined nature of our choices, as it is these ancient figures that lead us to our particular choices, thus to shaping our forms of fate. However, individuals also possess a personal ego, which is able to form a new personal fate from the potential possibilities offered by the familial unconscious. Fate analysis holds that our possible fate always contains certain compulsions, (due to familial ancestry, heredity, and genetic determination) but the free choice of the ego is always present, which can result in a freely chosen destiny. Fate analytic therapy aims to confront patients with the legacy of their ancestors, and acknowledging that they will be able to decide freely how they are going to handle this heritage, i.e. what new fate they want to create for themselves (Huth, 1978, Kürsteiner, 1987, Jüttner, 2003).

Freedom and compulsion in human fate

The question of fate had already occupied Szondi's interest in the early days of his career (Szondi, 1937). It then happened that a married couple visited his consultancy, and – years after their wedding – the wife presented symptoms which were most similar to those of her mother-in-law. The wife complained of insomnia, headache, fear of social situations, as well as neurotic obsessions of the urge to poison others. This case reminded Szondi so much of an elderly patient of his that after finding the notes of her case he was convinced that at that earlier date the widow had used almost the same words to describe her symptoms. The husband, present in the consultancy, recognized his mother in the elderly lady. Szondi thought that most of the contemporary doctors would have considered this a matter of sheer coincidence – that is, a man choosing a wife who happened to suffer from the same poisoning obsessions as his mother years earlier. However, Szondi himself held different views, and searched for the answer to the question of what could have led the husband in his choice of spouse. Szondi based his answer on genetics, and claimed that the husband and the wife were 'gene-relatives', whose fate was determined by their common 'inherited elements'. Supposedly, the disorder that manifested in both the mother and the wife was latently carried by the husband's hereditary features. Szondi claimed that it was these latent genes that unconsciously guided the husband in his partner-choice. Later the question was extended beyond the choice of a partner to that of what directs humans in their crucial existential choices? In other words, what leads us in our choices of partner-, friend-, profession-, illness- and form of death? In his research Szondi recorded several hundred genograms (family trees) and collected data on thousands of relatives in order to answer these questions with the help of genetics. Finally, he came to the conclusion that ancient drives determine our choices. The act of choice is perfectly conscious, however the cause of the choices lies hidden in the familial unconscious. Szondi claimed that in our crucial existential decisions we are all guided by this familial unconscious. The familial unconscious contains those unconscious drives that strive to return in the fate of the offspring, following such genetic laws as 'patterns and figures' (Hughes, 1992).

In his first fate analytical study (1937) Szondi elucidated the genetic background of partner-choice and described its types. It was then that he formulated the fate analytical

object-choice theory, which he called genotropism. According to the concept of genotropism, two individuals whose genetic material is similar, may carry latently reappearing hereditary tendencies, and will mutually attract each other. By 1944 Szondi had extended the choices that determine our fate beyond the choice of a partner to the choice of friends, illness and the form of death. Ancient drives lying hidden in the familial unconscious also play a role in guiding one's choices.

In his first approach – during his investigations in Hungary – Szondi chiefly studied the factors determining one's fate (family heredity, social and mental milieu), but already at that time he had established the concept of dirigible fatalism. According to the concept of dirigible fatalism, nature will precisely circumscribe one's individual sphere of fate, but within this genetically determined sphere we are free to consciously choose objects and ways of life which help us gratify our ancient drive needs in our individual ways. This is the reason why fate analysis uses the term dirigible fatalism. The ancient drives may be gratified both in socio-positive and socio-negative forms, i.e. in ways accepted or rejected by society. Szondi holds that the choice of profession is a domain where we are able to channel our drives in a positive manner. (Just to mention a few of the abundant examples: a man may become a firefighter in order to socialize his pyromaniac drives; or may chose to be a butcher, a surgeon or a dentist to socialize his aggressive drives. Szondi also describes all the humanized, socialized or sublimated ways of gratifying our drives.)

In summary, the first phase of fate analysis mainly focused on searching for factors determining fate. Szondi thinks that people's whole walk of life is impregnated by genetic factors, which determine the force of drives, the process of development, the direction of manifestation and shape individuals' fate. However, the functioning of the genes does not entirely determine our sphere of fate. Our biological attributes are in constant interface with the environment, and their mutual impact creates a flexible and dynamic system, which may result in a number of unique variations. Consequently, although the various manifestations are closely related to the ancestors, the role they play in life may be highly diverse because of the individual and societal aspects.

In his research period in Switzerland Szondi already placed the emphasis on the choice of fate in the concept of new anancology (Szondi, 1954, 1968). At this time Szondi's main questions involved whether individuals have only one single fate. If inheritance endows us with several possible fates are we really able to choose from them? Do people have a fate that can be chosen freely besides the fate determined by an inherited compulsion? If individuals are born with the possibility of more than one fate, how could those be made conceivable and conscious for them? In reply to these questions, Szondi elaborated the system of fate analytic therapy, which aims to help the patient to 'swap' the negative fate for a new, better one. From this point on, both in theory and therapeutic practice, Szondi distinguished and used the notions of a compulsive fate and a freely chosen fate. He claims that everyone's fate holds both compulsion and freedom. The building blocks of one's destiny are handed down by the ancestors, thus the building material we may have is predetermined. However, it depends on our free choice what individual fate we are able to formulate or integrate with these blocks. In other words, familial unconscious contains a collection of patterns for all the possible fate-figures, from which individuals are free to choose one pattern and figure, or, as in most of the cases, to create their unique destiny from various possibilities of fate. Szondi's new anancological concept accepts the choosing, decisive role of the ego in human fate, which, at the same time, provides freedom and places responsibility on individuals in directing their fate. In this way, personal destiny is a dialectical coexistence of compulsive fate and freely chosen fate. (See Figure 1.)

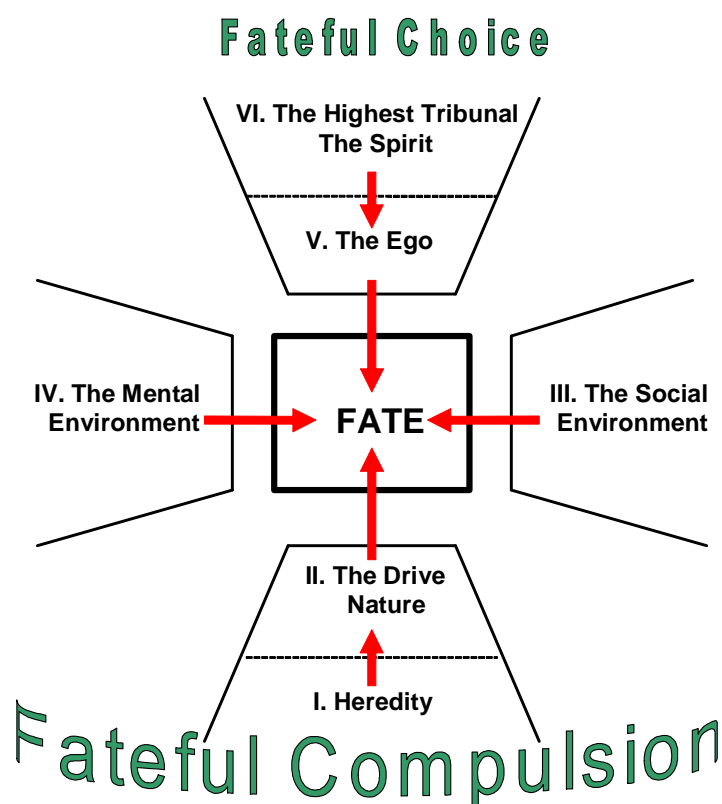


Figure 1. The scheme of dialectic anancology, Szondi 1954.

According to Szondi, in the life of the individual there is a definite plan. This plan comprises six forces. These are: 1. heredity, 2. the character of drives, 3. the social milieu (environment), 4. the mental milieu (environment), 5. the ego and 6. the spirit which is the highest tribunal of the fate. (Heredity, the character of drives, the social and mental milieu are the factors of compulsive fate, while the ego and the spirit are the factors of selective fate). The conception of a transformation of fate occurring in the course of time is based upon the fact that of these six fate-determining forces of the individual, at one time the compulsive power of heredity, drive nature, social or mental milieu are in the foreground, but at another time, the freedom of the ego and the spirit predominate choice. The fate of the individual is neither compulsively predetermined entirely by internal and external factors nor yet absolutely free (Szondi, 1954).

Thus, in Szondi's model of fate compulsive fate, is composed of the system of drives formed by genetic inheritance, the mental and the social milieu, while freely chosen fate consists of the decision-making (stand-taking) ego and the spirit. The system of drives built upon the genetic background creates the features of the personality, whereas the mental- (or world-view) and social environment is the very milieu the individual is born into. People are not free to choose these aspects, i.e. these are the compulsive components of one's fate. The factors of the freely chosen fate are the decision-making (stand-taking) ego and the spirit supporting it. The latter contain all the ideas that belong to the person's value system and enable the individual to experience supra-personal transcendence.

Thus, in fate analysis, besides heredity, the decision-making (stand-taking) ego and the spirit are also considered to be powers directing people's destiny. The ego bridges over drives and the spirit. The spirit is the transcendent instance that provides ancient drive forces with humane features and direction. The eight factors determining fate – together and opposing one

another – are in constant movement and change during our lifetime, thus their interaction is dialectically interpreted. The dialectics and constant dynamic change of fate factors force us always to be ‘underway’ and never to ‘stand still or arrive’. Using a theatrical metaphor, Szondi explains how fate changes on one’s stage of life, just as acts and scenes change on a revolving stage in the theatre.

Szondi’s fate analysis, in which he intended to study the genetics of the familial unconscious, may serve as a bridge between Freud’s personal unconscious and the collective unconscious described by Jung. Familial unconscious manifests itself in individuals’ fate choices by the inheritance of the ancestors’ aspirations and strivings. Szondi argues that the above three branches of depth psychology intend to elucidate the three different functioning modes of the unconscious by voicing the language of symptoms (Freud), symbols (Jung) and choices (Szondi). These ways of functioning all exist in the psyche in continuity and united globality (Szondi, 1955).

Fate analytical study of drives

In his study of drives Szondi elaborates on Freud’s ideas when defining drives as certain compulsions in the living organism, striving to restore a former state. Furthermore, Szondi states that from this follows the genetic origin of drives. He also diverges from Freud in rejecting classification of drives into life- and death drives and claims that the number of drives equals the number of genes determining drive-directed reactions. People with mental disorders actually suffer from drive disorders, or drive syndromes. Relying on the contemporary findings of genetics, he distinguishes four drives: the sexual, paroxysmal, schizoform (ego) and circular (contact) drives. Each drive consists of two opposing drive needs, and each drive need comprises two opposing drive tendencies. These 16 drive tendencies account for the dialectics and dynamics of our drive system. Szondi does not make a qualitative distinction between the drive needs of healthy and mentally disordered individuals. This means that we all have the same drive needs, the difference between people is only quantitative. To understand the concepts of drives, drive needs and drive tendencies, see Figure 2.

Sexual drive h ↔ s		Paroxysmal drive e ↔ hy		Ego-drive k ↔ p		Contact drive d ↔ m	
to love sy, or need for being loved	sadism	being good	to show oneself	introjection	inflation	seeking sy or sg	clinging to sy or sg
↕	↕	↕	↕	↕	↕	↕	↕
humanistic love towards mankind	masochism	being bad	to conceal oneself	negation	projection	sticking sy or sg	become detached
Eros	Thanatos	Etics	Morals	Ego-systole	Ego-dyastole	Anality	Orality

Figure 2. The drive system in fate analysis

The four psychopathologic drive cycles and their eight corresponding inherited mental disorders by Szondi as follows:

- I. Sexual drive: 1. homosexuality (h), 2. sadism (s),
- II. Paroxysmal drive: 3. epilepsy (e), 4. hysteria (hy),
- III. Schizoform drive or Ego-drive: 5. catatonia (k), 6. paranoia (p),
- IV. Circular or Contact drive: 7. depression (d), 8. mania (m).

Szondi developed his projective test in the years around 1935. The test consists of the photo portraits of 48 mentally ill patients suffering from 'drive syndromes'. The pictures are exposed in six series, each series displaying the portrait of a patient suffering from one of the above eight disorders. The patient is asked to choose from among them according to his/her preferences, likes and dislikes. Initially the test was called a 'genotest', as it aimed to detect the patient's drive structure. Adminstrating the test on ten occasions enables the researcher or the therapist to map both the patient's current psychic state and his/her permanent personality traits. The test is also suitable for the diagnosis of the character- and drive structure of normal individuals and patients with psychopathology.

Ego-functioning and personality in depth psychologies

The ego-concept of fate analysis differs from both that of psychoanalysis and analytical psychology. In Freudian psychoanalysis the ego develops from a part of the id and after its formation it functions as a mediator between the id and the outer world. Ego-functioning needs to harness the id, and to adjust the outer world to fit the desires of the id. In order to divert the unfeasible desires of the id, the ego develops various defense mechanisms, such as suppression, negation or sublimation. Besides the id and the ego, the third sphere of the psychic apparatus is the super-ego, which is formed through the internalization of parental requirements and prohibitions. The super-ego serves to prevent the realization of the drive tendencies disapproved by society, to direct the ego from rationality towards moral goals, to elevate the ego to the realm ideas. The three instances are in conflicts: instead of the pleasure principle of the id, the ego adheres to the reality principle, whereas the super-ego with its ideals directs the ego away from the reality principle towards irreality. A healthy personality is able to form a relative balance between the three instances of the psychic apparatus, thus ego-functioning is not destroyed by either id drives or the strivings of the super-ego. Beyond describing the general structure of the psychic apparatus, the unique personality can also be grasped in Freudian psychoanalysis. This unique personality is an individual pattern, based on the interaction of drive forces and defense mechanisms operated by the ego (Freud, 1940).

In Jungian analytical psychology personality is a complex notion, involving various partial complexes, and it has both conscious and unconscious aspects. Personality contains psychic structures as the ego, persona, shadow, anima and animus, and the *Selbst* (Self or the deep core). The ego is the center of the field of consciousness, and as it involves the empiric personality it is the subject of all conscious personal acts of the ego. As it could be seen earlier, the ego, or the ego-complex is only a part of the whole personality. The ego serves to maintain the functioning and the continuity of the personality; it is through the ego that we can experience identity. Testing reality is a further task for the ego, so that we are able to assess our inner processes, as well as the outside world. It is also the ego that perceives oppositions, conflicts between conscious and unconscious. The ego constantly strives to assimilate certain contents of mind, and attempts to elevate them into the experienced field of conscious functioning. Another important aspect of the Jungian psychology is the description of complexes, as becoming familiar with them will also lead individuals further on the way of acquiring and comprehending their unique personalities. Complexes are autonomous,

emotionally charged parts of the psyche, which can have positive, motivating effects; and negative ones, originating from an (often childhood-) psychic trauma and influencing later psychic development. Complexes are unconscious, thus making them conscious and affectively elaborating them helps the rearrangement of formerly engaged energies. This will result in the restoration of the psychic balance and the wholeness of the personality, as well as the reintegration of the dissociated parts of the psyche. In his analytical psychology Jung emphasizes the significance of individuation in the evolution of individual existence. In the individuation process the first step of psychic development is making the shadow conscious, then understanding the anima and animus follows, finally people need to reach their deep real self, and find the inner core of their psyche (Jung, et al. 1968).

In Szondi's fate analytical approach, the development of the personality is determined not only by genetic factors, but also by the mental- and social milieu, the stand-taking, decision making ego and its transcendent values. The term decision-making (stand-taking) ego is used by Szondi, and it denotes the function of our ego that makes decisions concerning ourselves. Szondi supposes that individual life as well as the development of the whole human race is based on a structure of oppositions. Such oppositions are the formerly discussed opposing drive needs, or counterparts such as conscious and unconscious, masculinity and femininity, body and mind, being awake and dreaming, the material and the transcendent world. The question remains, however, how, by whom, or by what means these oppositions are resolved. For the individual the ideal solution would be the integration of the oppositions, i.e. creating a new wholeness from the oppositions. This solution would be the most beneficial both for the individual and the society. If individuals try to eliminate the state of oppositions by preferring one pole and neglecting or suppressing the other, it is not only harmful for themselves but also undesirable for their environment, as they will pose a constant menace for society owing to their unsatisfied drives. (The counterparts of masculinity and femininity may be seen as an example, as choosing only one extreme and rejecting the other would not be beneficial. Men need to learn to be tender, and women proactive. People need to find the right balance of masculinity and femininity according to their gender roles, and they need to be able to integrate these opposing forces in their personalities.) The process of resolving oppositions is executed by a higher instance in the central control of the psyche, which Szondi calls *Pontifex Oppositorum*. Beyond the task of resolving oppositions, this Pontifex ego plays the role of assigning power and organization in the psyche, serving to bridge oppositions and create wholeness, the integrated personality. As discussed earlier, in his drive psychology Szondi describes four drives, each of them consisting of two opposing drive needs, which contain two opposing drive tendencies each. The dynamics of psychic life is based on this opposing nature of drive forces. It is the Pontifex ego that decides how the individual will integrate the drive structure constituted of oppositions, which drive needs it will gratify in a natural way, or which of them it will render to socialize or humanize (Szondi, 1956, 1980, 1984).

In Szondi's fate analytical psychology, the choice directed by the Pontifex ego is of utmost importance as it plays a significant role in the humanization of drives. The way of becoming a 'humane human being' is always open for people; however, they can only proceed on this way if – from the various possibilities of existence – they are able to make the humanized direction conscious, and then choose freely. In fate psychology it is the socialization, sublimation and humanization of drives that lead to the highest level of human development. As is manifest in his study of *Cain and Moses*, from these three important processes it is mainly the recognition and description of the role of humanization that we can attribute to Szondi (Szondi, 1969, 1973). For Szondi, humanism is an intellectual, spiritual movement striving for the spread of love and humane attitudes. According to fate analysis, humanistic people will choose the humane way from the opposing drives (humanistic and

non-humanistic manifestation); further, they are able 'to extend humanistic needs not only to themselves, their family members, their fellow-churchmen, social cast, race, and nation, but to all human 'objects' of the world' (Gyöngyösiné Kiss, 1995, 1999a,b, 2007). To reach this level of development individuals need the assistance of such ego-functions as transcendence, integration and participation. Transcendence enables individuals to step to a higher level of development; integration means the formation of a whole personality together with its conscious and unconscious aspects; and participation supports the ability of reintegration, which helps us to participate in human relationships, the world and the spirit.

Figure 3. details Szondi's ideas concerning the image of the human being in various psychological movements.

FUNCTION	0 participation, projection	I sublimation	II compensation	III self actualization, inflation	IV - V humanisation	
The developmental stages of becoming human					D liberator and humanisator	
HOMO				C individuator	elector	homo humanus
	0 participator	A repressor	B potentator			
The cronological order of the different directs of depth psychology	Imre Hermann, René Spitz	psychoanalysis (S. Freud)	individual psychology (A. Adler)	analitical psychology (C. G. Jung)	fate analysis (L. Szondi)	humanism

Figure 3. The road to becoming human (Szondi, 1963b)

Fate analytical therapy

Fate analytical therapy aims to enable patients to leave the compulsions of their fate and directs them toward a more freely chosen way of existence (Szondi, 1963a; Seidel, Jüttner, Borner, 2002). In patients' lives compulsive fate refers to the power and determining role of hereditary genetics, which may deteriorate the quality of their lives. The possibility of the socio-positive canalization of disturbing genetic factors may help individuals experience less suffering, and gratify morbid drive forces in a positive form. For the purposes of the fate analytical therapy Szondi used a curriculum vitae written by the patient, a detailed genogram or family tree (preferably checked by the family doctor), and a ten-profile Szondi test. At the beginning of the analysis the therapeutic techniques are similar to those applied in psychoanalysis, being based on free associations. Patients also talk about their dreams. The therapeutic process is modified only months later, when the therapy reaches the so-called 'gap-period' or 'hole-period'. In this period the patient's thoughts seem to be stuck, the chain of association from the depths is broken. A patient once described this situation as 'standing by a gap, not being able to proceed'. This halt or breakdown of the process, when associations also become superficial may even last for months. To overcome this difficulty Szondi modified the

association technique and introduced the so-called ‘hammer-beat association’ technique, which aims to break the patient’s opposition that causes the block. The technique means that the therapist will very rapidly – like a hammer-beat – expose phrases and sentences from the material of former associations and dreams that are supposedly connected to the current block, or may do the same with all the words of the entire material of the chain of associations. At this point the patient lying on the couch will give up his/her opposition and reproduces the particular symptoms (epileptic, paranoid, catatonic etc.) of his/her ill ancestors. The symptoms are presented and experienced by the complete consciousness of the patient, and this gives him/her a chance to face and experience the formerly hidden inheritance with clear consciousness. Szondi argues that as a result of this shocking effect the revolving stage of the psyche may be reversed, and the formerly hidden dimensions of the personality may be displayed. The hidden part of the personality may already be inferred from the earlier part of the analysis, when the patient shares with the therapist such ‘archaic dreams’ in which the protagonists are manifest carriers or conductors of a certain disorder. These ancient ill figures really exist in the patient’s portfolio of ancestry as ‘patterns of existence’. The main aim of the fate analyst is to understand the patients’ ‘conductor nature’, as it is the carrier of their fate neurosis. Fate analysis intends to disclose the latently present ill ancestors, which – according to genetic laws of heredity – strive to ‘reappear in descendants’ and negatively influence their fate.

In addition to the symptoms vividly experienced in the course of the analysis the family tree and results of the Szondi test also play an important role in confronting the patient with his/her latent tendencies.

The last phase of the therapy is the post-confrontation phase, in which the therapist and the patient try to find alternative forms of gratification for the formerly morbid drive forces that may be fit in the patient’s life plans, and are not threatening for society. Let us remind the reader here that in Szondi’s view mental disorders are actually drive syndromes, and it is the dominance of drive forces that leads to pathologies. Apart from the pathologic forms, drive forces may be gratified in several other ways, for example through the chosen profession, hobby or in different fields of interest. It is an important criterion for fate analysis that it can help patients whose ego-functions are intact, i.e. the ego is able to take a stand concerning heredity. Szondi’s fate analysis is an efficient therapeutic technique to cure transgenerational, hereditary neuroses or fate neuroses.

Szondi and modern biology

In the 1930’s Szondi based his fate psychology on the achievements of leading researchers of genetics (e.g. Brugger, Weinberg, Johanssen), so it was the Mendelian dominant-recessive genetic law that he based his concepts on the hereditary nature of ‘drive genes’ that provide the biological background of the personality. Szondi even bases the law of genotropism on ‘genesmanship’, i.e. he supposes that our significant choices are influenced by the effect of recessive genes functioning. Although Szondi occasionally exceeded the limits of his contemporary genetic knowledge, – e.g. when discussing the biological origin and function of drives – in general, he retains the views of genetics of his days. The question arises as to how we are to evaluate Szondi’s drive theory and the phenomenon of genotropism from the perspective of our present-day knowledge of genetics. As it is clear from many recent researches (e.g. the genetic and evolutionary approach to personality, personality theories based on neural functioning, modern temperament- and character theories etc.), the hereditary nature of certain personality traits has been confirmed. Nevertheless, the genetic cornerstones of Szondi’s theory seem to fail the test of modern scientific thinking. It is widely accepted, for

example, that personality traits are determined not only by one but several, supposedly hundreds or thousands of genes, which affect thinking and behaviour in interaction with environmental factors. In spite of this, Szondi's concept could be modified, if instead of allele pairs and Mendel's laws, the theory of polygenic inheritance were to be applied. The idea of originating the phenomenon of genotropism from recessive genes is also problematic from the perspective of present-day genetics. This problem, however, is not necessarily irresolvable either, as the question of latent genes can be approached in modern genetics, for example, in the phenomenon where after passing a certain threshold genes of lesser effect may determine a given behavior or disorder (Bereczkei, 1995; Bereczkei, Gyöngyösiné Kiss, 2001). In summary, revisiting and reassessing Szondi's theory in light of modern genetics is yet to be seen.

The significance of the Szondi test in comparing different nationalities

The Szondi test is a projective technique which uncovers the features of personality. As a projective technique, its purpose is to establish a testing situation where the subject is enabled to express his inner world without knowing what he really reveals. It follows that the test assesses the unconscious level of personality which is not accessible directly. Furthermore the Szondi projective test shows not only the basic structure of personality but the dynamic aspects of it as well, such as the accumulation and discharge of the drive-needs and need-tensions. Using the test it is possible to assess the cultural influence on personality as has been shown by several cross cultural studies (Yarritu, 1955, Szondi, 1972, Yamashita, 1999).

References

- Bereczkei, T. (1995). The Szondi's legacy: Innate dispositions influence our choices. A sociobiological reinterpretation of the Szondi-theory. *Szondiana, Zeitschrift für Tiefenpsychologie und Beiträge zur Schicksalsanalyse*, 15, 8-26.
- Bereczkei, T. & Gyöngyösiné Kiss, E. (2001). Sorsanalízis és genetika: családfakutatás, párválasztás, öröklődés. (Fate-analysis and Genetics: family-tree research, choice of a partner, heredity). *Magyar Pszichológiai Szemle*, 56, 71-90.
- Bürgi-Meyer, K. & Gyöngyösiné Kiss, E. (1994). Verzeichnis der Schriften von Leopold Szondi. *Szondiana, Zeitschrift für Tiefenpsychologie und Beiträge zur Schicksalsanalyse*, 14, 57-63.
- Bürgi-Meyer, K. (2000). *Leopold Szondi. Eine biographische Skizze*. Zürich: Szondi Verlag.
- Déri, S. (1949). *Introduction to the Szondi test: Theory and practice*. New York: Grune & Stratton.
- Freud, S. (1938/1940). *Abriss der Psychoanalyse*. Gesammelte Werke, Band 17. Frankfurt a.M: Fischer Verlag.
- Gyöngyösiné Kiss, E. (1995). Schicksal analysis and Jewish faith. *Szondiana, Zeitschrift für Tiefenpsychologie und Beiträge zur Schicksalsanalyse*, 14, 27-34.
- Gyöngyösiné Kiss, E. (1999a). *Szondi Lipót sorsanalízise* (The fate-analysis of Lipót Szondi). Ph.D. thesis, University of Pécs.
- Gyöngyösiné Kiss, E. (1999b). *Szondi Lipót*. (Lipót Szondi) Budapest: Új Mandátum Könyvkiadó.
- Gyöngyösiné Kiss, E. (2007). Mélylélektan és emberkép. (Depth Psychology and becoming human) In: B. Péley & Gy. Révész, Gy (Eds.), *Autonómia és identitás* (pp. 59-69). Pécs: Pro Pannonia Kiadói Alapítvány.

- Hargitai, R. (2007). The Szondi test: Theory and practice. *Empirical Text and Culture Research*, 3, 85-105.
- Hughes, R. A. (1992). *Return of the ancestors*. New York: Peter Lang.
- Huth, W. (1978). *Wahl und Schicksal. Voraussetzungen, Grundprinzipien und Kritik der Schicksalsanalyse von Leopold Szondi*. Bern: Verlag Hans Huber.
- Jung, C. G., von Franz, M.-L., Henderson, J. L., Jacobi, J., Jaffé, A. (1968). *Man and his symbols*. New York: Dell Publishing Group, Inc.
- Jüttner, F. (2003). *Schicksalsanalyse in Zusammenfassungen*. Zürich: Szondi Verlag.
- Kürsteiner, G. (1987). *Schwerpunkte der Schicksalsanalyse*. Berner Forum Tiefenpsychologie. Band 1. Bern: Kürsteiner, G.
- Seidel, P., Jüttner, F. & Borner, M. (2002). *Manual der schicksalsanalytischen Therapie*. Zürich: Szondi Verlag.
- Szondi, L. (1937). *Analysis of marriages. An attempt at a theory of choice in love*. The Hague: Nijhoff.
- Szondi, L. (1944). *Schicksalsanalyse*. Basel: Schwabe.
- Szondi, L. (1947). *Lehrbuch der Experimentelle Triebdiagnostik*. Bern: Verlag Hans Huber.
- Szondi, L. (1952). *Triebpathologie*. Bern: Verlag Hans Huber.
- Szondi, L. (1954). Mensch und Schicksal. Elemente einer dialektischen Schicksalswissenschaft (Anankologie), *Wissenschaft und Weltbild* 7, Wien: Verlag Harold, 15-34.
- Szondi, L. (1955). Die Sprachen des Unbewussten: Symptom, Symbol und Wahl. In : *Szondiana II. Beiheft zur Schweizerischen Zeitschrift für Psychologie und ihre Anwendungen*, Bern: Verlag Hans Huber.
- Szondi, L. (1956). *Ich-Analyse*. Bern: Verlag Hans Huber.
- Szondi, L., Moser, U., & Webb, M.W. (1959). *The Szondi test in diagnosis, prognosis and treatment*. Philadelphia & Montreal: Lippincott Company.
- Szondi, L. (1963a). *Schicksalsanalytische Therapie*. Bern: Verlag Hans Huber.
- Szondi, L. (1963b). Der Weg zur Menschwerdung. In: *Szondiana IV. Beiheft zur Schweizerischen Zeitschrift für Psychologie und ihre Anwendungen*, Bern: Verlag Hans Huber.
- Szondi, L. (1968). *Freiheit und Zwang im Schicksal des Einzelnen*. Bern: Verlag Hans Huber.
- Szondi, L. (1969). *Kain. Gestalten des Bösen*. Bern: Verlag Hans Huber.
- Szondi, L. (1973). *Moses. Antwort auf Kain*. Bern: Verlag Hans Huber.
- Szondi, L. (1980). *Die Triebentmischten*. Bern: Verlag Hans Huber.
- Szondi, L. (1984). *Integration der Triebe. Die Triebvermischten*. Bern: Verlag Hans Huber.
- Yamashita, T. (1999). Szondi test responses in Japanese juvenile delinquents. *14th International Szondi Conference*, 15-17 July 1999, Louvain-la Neuve, Belgium.
- Yarritu, S. (1953/1955). Validierung des Szondi-testes durch eine Gruppenuntersuchung von 2352 Fällen. *Szondiana. Zeitschrift für Tiefenpsychologie und Beiträge zur Schicksalsanalyse*, 2, 65-75.

Comparing the Szondi Test results of Hungarian and Portuguese community samples

Bruno Gonçalves*, Ana Ferreira*, Mátyás Káplár,
Enikő Gyöngyösiné Kiss****

*Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade,
1649-013 Lisboa, Portugal

**University of Pécs, Institute of Psychology, Department of Personality -,
Developmental and Clinical Psychology, H-7624 Pécs, 6. Ifjúság útja, Hungary

Abstract

This study makes a comparison between two nations' Szondi test profiles. The aim of this study was to compare the Szondi Test results of a Hungarian community sample (n=126) with those of a Portuguese community sample (n = 175). Both samples were heterogeneous convenience samples. Results show a general agreement (with some minor differences) in 5 factors: h, e, hy, k, and d. There are very clear differences in s and p factors, where s- and p+ are the most frequent results in the Hungarian sample vs. s+ and p0 or p- in the Portuguese sample. Differences in the m factor are significant but less important: m+ is the most frequent result in both samples but it attains a relatively higher frequency in the Portuguese sample, where it is also more frequently loaded (+!!). The differences in the samples' educational level contribute to the differences in p factor but do not explain all the observed differences. Age and gender seem to have only a slight influence on the results.

Keywords: Szondi projective test, comparison of a Portuguese and Hungarian sample

Introduction

The Szondi test is a projective technique. The test material consists of 48 photos representing the faces of mental patients. The subject is asked to tell which photos he/her likes the most and which photos he/her dislikes the most (Szondi, 1947; Deri, 1949). The readers of the journal can read about Lipót (Leopold) Szondi's fate-analytical theory in this edition by Enikő Gyöngyösiné Kiss and about the Szondi personality test in the previous edition by Rita Hargitai (2007). Szondi applied his test to a large sample of 1000 adults from the Hungarian general population. The most frequent responses in this sample defined a very similar profile to what Szondi had previously described as the profile of the "man in the street" (*Alltags-mensch*). One important feature of this profile is the disciplined ego (Sch --). In the first edition of his *Experimentelle Triebdiagnostik* (1947) Szondi stressed that this type of ego was found most frequently in manual workers from lower social classes and that it was rare among "intellectuals". Reference to social classes was omitted in the 2nd and 3rd editions of this book.

Soto-Yarritu (1955) studied a sample of 750 adults from the Spanish (Navarre) general

population. Overall, as can be seen in Table 4, his results were consistent with those obtained by Szondi. However, there were also some clear differences, namely in the C vector results. In the Navarrese the tendency to cling to the object in a more or less rigid way (C+) seem to be much more frequent than in Hungarians. Szondi (1947) attributed this difference to cultural characteristics, namely to the strong attachment of the Navarrese to the Catholic Church.

These kind of studies with large samples from the general population have not been repeated since. However, other studies with smaller samples, namely A. Pochet's study (data published by Satassart, 1999) with a sample of 200 Italians, yielded results in the C vector similar to those obtained by Soto-Yarritu. In the other vectors, results of these modern studies tend to confirm Szondi and Soto-Yarritu's results but they also show some differences, namely in the p factor, where the p+ tends to be much more frequent than p-, exactly the opposite of what was found by Szondi and Soto-Yarritu. This implies that the "disciplined" ego becomes relatively rare (7%). Gonçalves (1999) suggested that this difference could be related to the subjects' social level. Unfortunately, data on the subjects' educational or social level are generally not given on all these studies with the Szondi test. On the other hand, as we have no modern studies with the Hungarian population, we do not know if Szondi's results come from stable cultural characteristics or whether they were influenced by historic circumstances or by the specific social characteristics of the initial sample studied by Szondi.

The aim of this study is to compare the results of a Hungarian and Portuguese community sample taking into account the possible effect of differences in age, educational level, work status and sex between the two samples.

Initial data of this comparison were presented by Gonçalves and Gyöngyösiné Kiss (2008), and Káplár, Gyöngyösiné Kiss, and Gonçalves (2008).

Method

Participants

Both the Hungarian (n = 126) and the Portuguese (n = 176) samples were heterogeneous convenience samples collected in non-clinical settings. Participants who volunteered aged between 17 and 65 years were accepted. Table 1 presents the main characteristics of the samples.

Table 1: Characteristics of the samples

		Hungarian (n = 126)	Portuguese (n = 176)
Age	(M ± SD)	27.4(± 8.9) years	30.4(± 10.3) years
	[Max-min]	[17- 60 years]	[17 - 63 years]
Sex	men	37.3%	57.4%
	women	62.7%	42.6 %
Educational level	<12 years	1.6 %	29.6 %
	≥12 years	98.5 %	49.4 %
Work status	univ. student	56.4 %	21.0 %
	worker	43.6 %	79.0%

The age mean of the Hungarian sample is significantly lower than in the Portuguese sample

(t-test = -2.65, $p = .008$). The percentage of women in the Hungarian sample is higher than in the Portuguese sample (chi-square = 11.85, $p = .0006$). There were also very clear differences in the educational level: almost all the subjects in the Hungarian sample had at least 12 years of schooling. In the Portuguese sample almost 30% of the subjects had less than 12 years of schooling. Finally, the percentage of university students was higher in the Hungarian sample.

Measures

In a clinical setting, the Szondi test normally includes 10 different profiles, obtained on 10 different days. However, for research purposes, a shorter, 5-profile form has often been used. All the subjects in the Hungarian sample had a 10-profile Szondi test. Most of the subjects in the Portuguese sample had only a 5-profile test (in which case results were extrapolated to 10 profiles). The total number of profiles collected was 1156 in the Hungarian sample and 1017 in the Portuguese sample. Data on the subjects' sex, age, profession and education level were also obtained.

Results

The frequency of the four possible responses (0, +, -, \pm) are presented for each factor and for both samples. Results presented in Table 2 are mean frequencies in %. Results for + and - responses include both loaded and unloaded responses. Results for loaded responses are also presented separately as they have a special clinical meaning and in some factors they are very frequent. Figure 1 presents the results in graph form.

Table 2: Frequencies (in %) of factorial responses in both samples. (*Mann-Whitney U test)

	Hung N=126	Port N=175	p – level*		Hung N=126	Port N=175	p – level*
h 0	28.6	24.6	.176	k 0	16.3	17.8	.917
h +	38.5	44.1	.276	k +	9.6	10.0	.086
h -	20.8	21.0	.221	k -	51.0	53.4	.555
h \pm	12.1	10.3	.130	k \pm	23.1	18.8	.033
h +!	4.2	13.0	.005	k +!	0.0	.4	.088
h - !	2.3	4.1	.723	k - !	5.6	6.8	.965
s 0	15.9	14.9	.580	p 0	28.5	34.4	.105
s +	22.8	45.7	.000	p +	61.3	27.1	.000
s -	38.1	21.7	.000	p -	6.6	31.9	.000
s \pm	23.2	17.7	.132	p \pm	3.6	6.6	.114
s +!	1.7	6.6	.014	p +!	14.9	4.3	.000
s - !	8.8	5.9	.004	p - !	.1	4.6	.000
e 0	31.7	34.1	.273	d 0	41.1	34.3	.023
e +	25.2	27.8	.278	d +	21.7	17.5	.003
e -	30.6	26.8	.385	d -	31.7	42.2	.011
e \pm	12.5	11.3	.469	d \pm	5.5	6.1	.293
e +!	1.1	1.2	.522	d +!	.9	2.5	.676
e - !	1.3	1.7	.652	d - !	2.5	5.0	.423
hy 0	10.7	16.9	.168	m 0	12.7	9.0	.003
hy +	4.1	3.9	.141	m +	66.4	77.5	.001
hy -	69.7	70.7	.576	m -	9.6	5.3	.002
hy \pm	15.5	8.5	.007	m \pm	11.4	8.3	.005
hy +!	.0	.2	.229	m +!	24.0	37.8	.001
hy - !	13.0	17.5	.737	m - !	.3	.6	.975

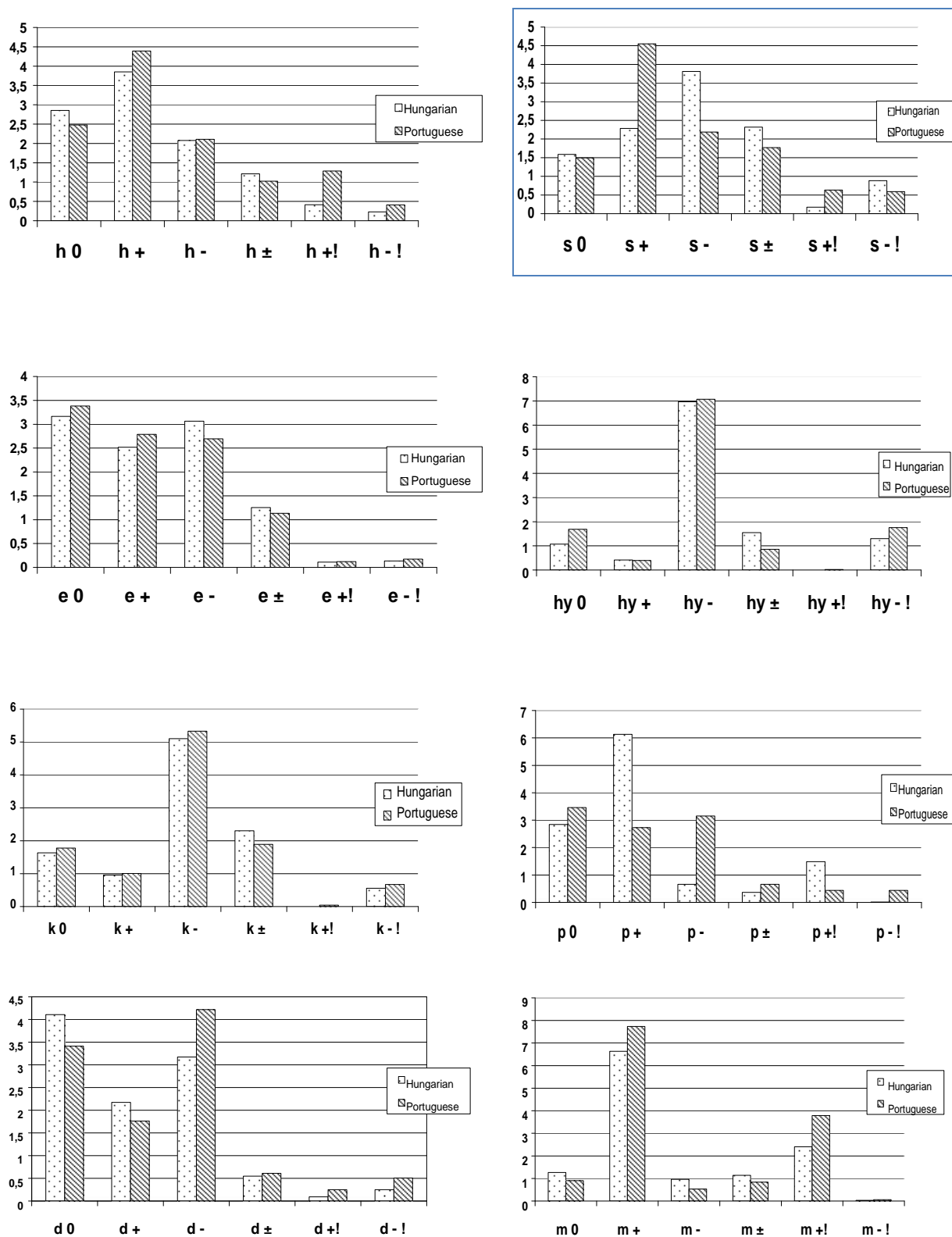


Figure 1: Frequency of different responses in each factor

In order to establish whether the observed differences could be attributed to differences in sex, education level or age between the Hungarian and the Portuguese samples, several separate log-linear regression analyses were conducted. The independent variables were sex, education (less than 12 years vs. 12 or more years of schooling), work status (student vs. worker), age

and nationality (Hungarian vs. Portuguese). The dependent variables were all the factorial variables presenting a statistically significant difference when the two samples were compared. These variables were dichotomized according to the median value (observed value under or equal vs. above the median value). For most of the variables the median value (MD) was zero. The frequency of k_{\pm} had a MD=1, the frequencies of s_{+} , s_{-} and $m_{+!}$ had a MD=2, and the frequencies of d_0 and d_{-} had a MD=3. Other studied variables with a median value above 0 were: p_{+} (MD=3.5) and m_{+} (MD=8.5). Data were analyzed by the statistical software SPSS 16.0 (SPSSInc., Chicago, IL., USA). Results are presented in Table 3.

Table 3: Logistic regression analysis: odds ratio for independent variables with a significant effect (Wald statistic).

Independent variables					Dependent variable
Portuguese /Hungarian	Age	Sex	Student/ worker	Education high/low	
-	-	-	-	.28***	h +!
4.22***	-	-	2.18*	-	s +
.21***	-	-	.41*	-	s -
.37*	-	-	-	-	s +!
3.64***	-	-	-	-	s -!
.31***	-	-	-	.47*	hy \pm
-	-	-	-	-	k \pm
.24***	-	-	-	3.44**	p +
4.38***	-	-	-	.20***	p -
.34***	-	-	-	4.11*	p +!
26.03**	-	-	-	.23**	p -!
.61*	-	-	-	.43*	d 0
.30***	-	-	-	-	d +
-	-	-	2.49*	-	d -
.29***	-	-	-	.35**	m 0
1.87*	-	-	-	-	m +
.34**	1.04*	-	-	-	m -
.46**	-	-	-	-	m \pm
2.44***	-	-	-	-	m +!

* $p < .05$ ** $p < .01$ *** $p < .001$

To compare the structure of our two samples we used a model of Agglomerative Hierarchical Cluster Analysis (AHCA) using the Chi-squared measure and the complete linkage clustering for each one, considering the 48 variables (Gordon, 1981). The software used to compare the two AHCA results was implemented in S-Plus by Lorga (2004). To compare the obtained dendrograms (not presented here) we used the Spearman correlation coefficient between the ultrametric similarity matrices. The obtained value ($r_s = .272$) shows a low but highly significant ($p < .001$) correlation between the hierarchy of partitions obtained with the 48 variables.

The similarity between the two samples may be seen when we compare the classes of variables that are formed at the best level, which is the penultimate where two classes are formed in each sample. Table 4 compares the obtained variable clusters in the two samples at this level.

Table 4: Comparison of clusters at the penultimate level in both samples

		Hungarian sample	
		Cluster 1	Cluster 2
Portuguese sample	Cluster 1	h- h± h-! s+! s-! e± e+! e-! hy+ k0 k+ k+! k-! p+! p± d± d+! d-! m0 m- m± m-!	h0 s0 s- s± e0 hy± hy+! k± p+ d+
	Cluster 2	h+! hy0 p- p-!	h+ s+ e+ e- hy- hy-! p0 k- d0 d- m+ m+!

To compare our results with those obtained by Szondi, Soto-Yarritu and Pochet, we used the frequencies of the vectorial constellations (pairs of responses in the two factors in each vector) and not the results for each separate factor. The three most frequent constellations in each vector are presented in Table 5. The synthetic global profile based on vectorial constellations includes the most frequent constellation in each vector and all constellations whose frequency is higher than half the frequency of the most frequent constellation and above 10%. Constellations printed on a grey background are not to be included in the profile.

Table 5: The most frequent vectorial constellations in the general population

	S		%	P		%	Sch		%	C		%
Szondi (n=1000)	+	+	26	+	-	23	-	-	32	0	+	18
	+	0	18	0	-	16	-	+	9	+	-	14
	+	-	13	-	-	8	0	-*	8	-	+*	10
Soto-Yarritu (n=750)	+	+	36	+	-	28	-	-	37	0	+	27
	+	0	14	0	-	21	0	-	14	-	+	27
	+	-	13	-	-	14	-	0	12	+	+	10
Pochet (n=200)	+	-	30	+	-	23	-	+	27	0	+	25
	+	0	12	0	-	17	-	0	10	-	+	13
	+	±*	8	-	-	17	0	-	8	0	±	12
Hungarian (n=126)	+	-	14	0	-	22	-	+	32	0	+	28
	+	0	12	-	-	20	-	0	15	-	+	23
	+	±	11	+	-	20	0	-	8	+	+	10
Portuguese (n=175)	+	+	22	0	-	26	-	0	19	-	+	37
	0	+	11	+	-	21	-	-	17	0	+	27
	-	+	8	-	-	16	-	+	14	+	+	10

*Other constellations have the same frequency.

Discussion

Results show a general consistency (with some minor differences) in 5 factors: h, e, hy, k and d. There are, on the contrary, very clear differences between the two samples in 3 factors: s, p and m. We will briefly comment on the results in each factor.

In the **h factor** there seems to be a general consistency. The two most common responses are the same in both samples. Loaded h+ is significantly more frequent in the Portuguese sample. However, the logistic regression analysis suggests that this difference results from differences in the educational level between the two samples.

In the **s factor** there seems to be a clear contrast between the two samples: s + is significantly more frequent in the Portuguese sample and s - is significantly more frequent in the Hungarian sample. The differences are also statistically significant when we consider only the loaded responses (s+! and s-!). Logistic regression analysis shows that s+ tends to be more frequent and s- less frequent in students, so the fact that there are more students in the Hungarian sample probably tends to attenuate the observed difference.

In the **e factor** there are no statistically significant differences between the two samples.

In the **hy factor** there seems to be a general consistency. Minus hy is the most common response in both samples. However, hy \pm is significantly more frequent in the Hungarian sample. Logistic regression analysis shows that this response tends to be less frequent in subjects with a higher education level, so the difference in education level between the two samples probably tends to attenuate the observed difference.

In the **k factor** there seems to be a general consistency, and results are similar to those obtained in the hy factor. Minus k is the most common response in both samples, but k \pm is more frequent in the Hungarian sample. However, the statistical significance of the observed difference in the frequency of k \pm is not as high as the correspondent value for hy \pm .

In the **p factor** there seems to be a very clear contrast between the two samples, since p + is significantly more frequent in the Hungarian sample and the opposite response, p -, is significantly more frequent in the Portuguese sample. The differences are also statistically significant when we consider only the loaded responses (p+! and p-!). Logistic regression analysis confirms that these differences are partially due to the differences in education level. As far as p - is concerned, the effect of educational level seems to be as important as nationality. However, even when we take into account the effect of educational level, there is still a statistically significant effect of nationality.

In the **d factor**, there seems to be a general consistency. The two most common responses (d- and d0) are the same in both samples (although not in the same order). On the other hand d0 and d+ are significantly more frequent in the Hungarian sample and d- is significantly more frequent in the Portuguese sample. Logistic regression analysis shows that d0 tends to be less frequent in subjects with a higher education level and d- tends to be more frequent in students, so the difference in education level between the two samples and the fact that there are more students in the Hungarian sample, both probably tend to attenuate the observed difference.

In the **m factor**, there seems to be a general consistency. The two most common responses (m + and m 0) are the same in both samples. However, there are also several significant differences: m + and m + ! are significantly more frequent in the Portuguese sample, while m 0, m - and m \pm are significantly more frequent in the Hungarian sample.

The comparative analysis of the global profiles from both samples suggests that Hungarians are more passive (s-, s-!) and idealistic (p+). The Portuguese seem to be more active or aggressive (s+, s+!), yearning for a participative link (p-) and in fear of losing their attachment objects (m+!).

The cluster analysis performed on both samples shows some interesting similarities. We will focus on the variables that are classed in the same cluster in both samples (Table 4). In Cluster 1 we find signs of problematic or immature contact ($m-$, $m-!$, $m\pm$, $m0$) related to loaded reactions in almost all the other factors, suggesting the danger of imbalanced drive dynamics. Some of these reactions ($m0$, $m-$, $m\pm$, $s-!$) are significantly more frequent in the Hungarian sample but we cannot say that they characterize the global profile of this same sample. In Cluster 2 we find almost all the reactions that can be included in what we could call a relatively stable neurotic profile with some attachment anxiety which is, in fact, the global profile of the Portuguese sample (see Table 5).

The fact that these variables are grouped in the same way in both samples suggests that they can also be interpreted in the same way in both samples. Interpretation can be more complex when the reactions are grouped in a different way. Such is the case, with the $p+$ and $p-$ reactions. In the Portuguese sample the $p+$ reaction is included in Cluster 1, suggesting a psychically imbalanced context, while in the Hungarian sample it is included in Cluster 2. The opposite can be observed as far as reaction $p-$ is concerned. In these cases a more detailed analysis of the global profiles including these reactions would be necessary.

The vectorial profiles (Table 5) confirm the above mentioned differences at the factorial level and allow for a quick comparison with results from previous studies using community samples.

In the **S and Sch vectors**, there are clear differences between the Hungarian and the Portuguese samples, resulting from the above mentioned factorial differences in the s and p factors. The results of the Portuguese sample are closer to those obtained by Szondi or Soto-Yarritu. The results of the Hungarian sample are very similar to those obtained by Pochet with an Italian sample from the region of Padova.

In the **C vector**, both the Portuguese and the Hungarian sample results are similar to those obtained by Soto-Yarritu and different from Szondi's initial results.

In the **P vector** there are no major differences between all the compared community samples.

There is no clear explanation for these results. Similarities between the Portuguese sample and the Spanish sample may result from shared cultural features. Northern Italy and Hungary are not very far away and both regions may also share some cultural features. However, there is no explanation for the fact that the results from the Portuguese sample in the **S** and **Sch** vectors seem to be closer to those obtained by Szondi. As far as the **Sch** vector is concerned, this similarity could result from the fact that the Portuguese sample has more subjects with a low educational level than the Hungarian sample.

One interesting issue for further research would be to find if there are differences related to educational and social level among the Hungarian population as have been found among the Portuguese population.

References

- Deri, S. (1949). *Introduction to the Szondi test. Theory and Practice*. New York : Grune and Stratton.
- Gonçalves, B. (1999). Tendances projectives et inflatives dans la population générale. *Szondiana*, 19, 2, 91-103.
- Gonçalves, B., Gyöngyösiné Kiss, E. (2008). Comparing test results from Hungarian and Portuguese community samples. 18. *Szondi International Society Congress, Lisbon, Portugal, 2008. 07. 10 - 07. 12.*
- Gordon, A. D. (1981). *Classification*. New York: Chapman and Hall.

- Gyöngyösiné Kiss Enikő: *Personality and the familial unconscious in Szondi's fate-analysis*. Empirical Text and Cultural Research, (present number).
- Hargitai, R. (2007). The Szondi test: theory and practice. *Empirical Text and Culture Research* 3, 85-105.
- Lorga, A. S. (2004). *Tratamento de Dados Omissos e Métodos de Imputação em Classificação*. Unpublished Doctoral dissertation. Universidade Técnica de Lisboa, Portugal.
- Káplár, M., Gyöngyösiné Kiss, E., Gonçalves, B. (2008): Az átlagprofil vizsgálatá hazai és kulturközi elemzésben. (Examination of the average profiles in national and cross cultural analysis) *XVII. Országos Pszichológiai Nagygyűlés*, (17th National Psychological Conference of the Hungarian Psychological Association) Nyíregyháza, Hungary, 2008. május 22-24. Abstract book, 232.
- Stassart, M. (1999). L'épilepsie essentielle aux tests de Szondi et de Rorschach. *15th Colloque de la SIS* [Abstract book]. Louvain-la Neuve, Belgium.
- Szondi, L. (1947). *Experimentelle Triebdiagnostik, Tiefenpsychologische Diagnostik im Dienste der Psychopathologie-, Kriminal- und Berufspsychologie, Charakterologie und Pädagogik*. Bern: Verlag Hans Huber.
- Yarritu, S. (1955). Validierung des Szondi-testes durch eine Gruppenuntersuchung von 2352 Fällen. *Szondiana. Zeitschrift für Tiefenpsychologie und Beiträge zur Schicksalsanalyse* 2, 65-75.

Hungarian ratings for the International Affective Picture System (IAPS): A cross-cultural comparison¹

Anita Deák *, Laura Csenki **, György Révész *

*University of Pécs, Institute of Psychology, 6. Ifjúság útja, Pécs, Hungary H-7624

**Semmelweis University, 1st Department of Pediatrics, Child and Adolescent Mental Health Unit, Budapest, Hungary

Abstract

Emotions serve as universal adaptations to social life. Social interactions, however, are implemented in a cultural context. In this experiment we examined how cultural context influences ratings in response to emotionally evocative stimuli. 187 Hungarian university students rated valence, arousal and dominance of pictures from the International Affective Picture System (IAPS). The affective ratings of the Hungarian sample correlated with the normative North-American results: 0.95 for valence, 0.72 for arousal and 0.69 for dominance dimension. Means of valence and arousal ratings did not differ significantly between groups. A significant difference was found in the ratings of dominance. Distribution of pictures in the valence-arousal affective space was similar in both samples. These results have underlined the biphasic motivational-emotional model of appetite and aversion that seems to be universal across cultures.

Keywords: emotion, cross-cultural studies, sex differences, IAPS, valence, arousal, dominance, appetite, aversion

Introduction

Emotions are complex phenomena, just like the theoretical assumptions and empirical studies about them. Emotional behaviour is determined by two main factors: one, reactions which involve biological mechanisms (such as facial expressions, gestures, cortical brain activity, subcortical brain activity, and hormonal changes); and, two, cultural and social norms, traditions and beliefs which regulate how to express or suppress affective states.

Certain evolutionary approaches to emotions and psychophysiological theories (James, 1884; Cannon, 1927) emphasize the biological aspect of emotions. These theories focus on emotional expressions which are universal across cultures, show similarities across species, occur automatically and function as communicative signals.

The evolutionary adaptive value of emotions has been discussed since the 19th century (Darwin, 1872/1963; Nesse, 1990; Damasio, 2000; Öhman, Flykt & Lundqvist, 2000; Pank-

¹ The authors thank Richard Sherry for his suggestions for this paper. Correspondence should be addressed to deakanita@btk.pte.hu.

sepp, 2006). They facilitate phylogenetic responses to environmental demands, change cognitive activities, serve interpersonal functions, and regulate social living (Clark & Watson, 1994; Frijda, 1994; Levenson, 1994; Scherer, 1994; Mesquita & Leu, 2007). Positive emotions have had the important function of strengthening social bonds or reinforcing ongoing behaviour, while negative affective states signal disturbances in social relationships.

The impact of culture on emotions, social constructivism emphasises stating that dynamic interactions between the individual and his environment shape emotions. Mesquita and Leu (2007) suggest that different cultures can have different representations about emotions. Subjective experiences and action readiness are influenced by cultural and social norms. Cultural models of a sociocultural context include cultural meanings and prescribe how to regulate feelings and appraise events.

Several studies have compared the cultural models of individualistic and collectivistic populations. However, other terms are also in use such as independent vs. interdependent context. The North-American context is often mentioned as independent, while context in East-Asia is considered to be interdependent. Cultural differences are expected in the level of subjective experience and appraisal across different sociocultural contexts.

In an experiment, subjects had to rate their affective state five times a day. They also had to indicate whether they were alone, with a friend, with their romantic partner, with a stranger, with family members or with a colleague. Ratings of subjective experience in an interdependent context (Japanese and Hispanic subjects) depended on the presence of other people in the situation. No difference was found in the independent context (American subjects with European roots) (Mesquita & Leu, 2007).

Cultural differences can be found not only in self-reported emotions but also in recognition of others' emotions. Japanese and American subjects were shown a cartoon with a figure surrounded by four other figures. They had to tell how the central figure would feel in that situation. In some cases, the facial expression of the central figure was the same as the others, in other cases they expressed different emotions. American subjects ignored other figures' emotional states and paid attention to the central figure. Japanese subjects took into account everybody's expression. For example, they rated a higher level of anger when an angry face was seen on one of the other four people's face. They also attributed a higher level of happiness to an angry central figure if the other four people showed happiness. These results indicate that the meaning of a social context is determined by the affective state of all participants for Japanese people. An emotional situation is assessed from other people's perspectives in an interdependent context. This is in contrast to an independent context where emotional meaning is seen from the person's own perspective (Mesquita & Leu, 2007).

The above mentioned social constructivist approach and appraisal theories imply that norms, traditions, and meanings of a social situation are culturally determined. Some components of emotions, however, are formed by biological factors.

Both biological and cultural components are integrated in the biocultural model of Levenson, Soto and Paul (2007). These authors suggest that cultural influences vary depending on which aspect of emotion is being considered. The strongest influence can be detected from self-reported emotional experience, somewhat weaker influence on emotional facial expression, and relatively minor influence on autonomic nervous system response.

Several methods can be used to study cross-cultural differences in emotion production and emotion recognition. Self-reports, measurements of physiological components, and emotion identification tasks are the most common instruments.

It is worth mentioning that translating emotional words from one language to another can be problematic in some cases. For example, the Polish language does not include any equivalents for the English word 'disgust' (Wierzbicka, 1986 in Levenson et al., 2007). Translational problems can be solved by giving a description about the emotion instead of

naming the category - for example, we can refer to fear as the experience that one can feel when he thinks that something bad will happen.

Scientists have argued about the nature of emotions in the sense that emotional experiences can be described by discrete categories and/or several dimensions. A fruitful approach to define emotional experiences is in terms of discrete emotions. Darwin and his followers have described several distinct emotional states, each of which seems to be accompanied by culturally independent distinctive expression. For example, smiling and wide lips are signals of happiness; a frozen stare, a pale face, coldness and erected hair show fear.

A common experimental procedure uses discrete emotional versus neutral faces as social signals of one's affective state and focuses on the detection of these stimuli (Baron-Cohen, Wheelwright, & Jolliffe, 1997; Martinez & Benavente, 1998; Baron-Cohen, Wheelwright, Hill, Raste & Plumb, 2001; Ekman, Friesen & Hager, 2002). An emotional face in a social context can inform about the affective state of the experiencer. However, recent studies not only focus on intrapersonal reactions and subjective feelings only, but they are also interested in the neural background (Breiter et al., 1996; Morris, Öhman & Dolan, 1998; Fitzgerald, Angstadt, Jelsone, Nathan & Phan, 2006; Suslow et al., 2006). A cross-culturally relevant topic is the brain activation occurring in response to the face of a person from the same or different race.

Much of the research brings into focus one's reactions to emotionally evocative stimuli (not only faces but scenes from several semantic categories)². In this case the perceiver can introspectively report about his feelings. In Plutchik's model (1994) not only discrete emotional states exist but it is possible to conceptualize emotions in terms of dimensions as well.

The history of a dimensional approach began with the work of Wilhelm Wundt. He proposed that affective meaning of a stimulus can be described by three basic dimensions: pleasant-unpleasant, excitation-inhibition and strain-relaxation (Wundt, 1897).

Later, Osgood (1962) found a similar dimensional structure when subjects were asked to rate stimuli on bipolar scales. He identified them as evaluation (good-bad, pleasant-unpleasant, positive-negative), potency (strong-weak, heavy-light, hard-soft) and activity (fast-slow, active-passive, excitable-calm).

In the 1970s, Mehrabian and Russel developed the Semantic Differential Scale for assessing stimuli. Their instrument consisted of 18 bipolar adjective pairs. Ratings were assessed in a three-dimensional space determined by pleasure, arousal and valence. Mehrabian and Russel's verbal method is difficult to use across cultures because of language and translation problems (Bradley & Lang, 2007b).

Followers of the dimensional approach agree that emotional events can be good or bad, appetitive or aversive, positive or negative, pleasant or unpleasant. This assumption has lead to biphasic motivational-emotional models (Bradley & Lang, 2007b). Biphasic theories assume that emotional signals refer to behavioural actions such as approach and avoidance. These actions are motivation-based and culturally independent.

Bradley and Lang developed a picture-oriented instrument called the Self-Assessment Manikin (SAM) (Bradley & Lang, 1994). SAM is an easy non-verbal method to measure a person's affective reaction to stimuli on three affective dimensions. Five small figures represent the scales (see Figure 1.). Pleasure/valence dimension ranges from a smiling, happy figure to a frowning, unhappy one; arousal dimension ranges from an excited to a relaxed

² Another interesting question is the psychological mechanism of detecting emotionally-evocative stimuli in different modalities. Several stimulus sets (visual, acoustic, olfactory) have been developed to induce emotions under controlled laboratory conditions (Bensafi, Rouby, Farget, Bertrand, Vigouroux, & Holley, 2002; Bensafi, Rouby, Farget, Vigouroux, & Holley, 2002; Bensafi, Rouby, Farget, Bertrand, Vigouroux, & Holley, 2003; Bradley & Lang, 2007a).

figure; the size of SAM figure changes on the dominance dimension, representing the subject's control over the affective state.

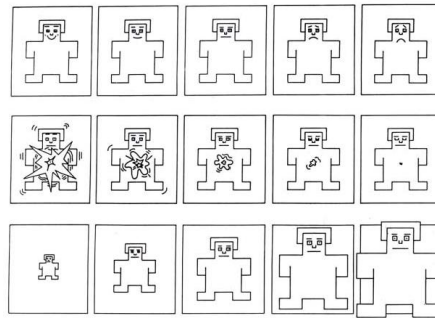


Figure 1. Three affective dimensions represented by SAM figures (Lang et al., 2005)

A set of standardized visual stimuli and a language-independent rating procedure make it possible to compare results from different laboratories, replicate special studies or examine cross-cultural issues. SAM ratings are comparable between males and females (Bradley, Lang & Cuthbert, 2005), different age groups (Backs, Silva & Han, 2005) and different cultures (Verschuere, Crombez & Koster, 2001; Ribeiro, Pompeia & Bueno, 2005).

A cross-cultural comparison was done involving North-American, German, Italian and Swedish samples (Bradley & Lang, 2007c). Swedish participants generally assigned lower arousal ratings to the pictures compared to US and German data, indicating calmer emotional reactions, whereas the Italians rated pictures as significantly more arousing. These data tend to support the general cultural stereotypes that exist for these countries.

The primary purpose of present study was to compare the affective ratings of IAPS pictures between Hungarian university students and the US norms published by Bradley, Lang and Cuthbert (2005). We investigated whether the normative ratings from the original US sample could be relevant in Eastern-European circumstances. Because little is known about specificities of this region, this research is highly relevant. Since the nature of emotions is determined by biological components as well as social impacts, the question of concern is whether a universal framework of emotions exists or cultural influences are represented more strongly in our study with IAPS.

Method

Participants

187 Hungarian students (146 women, 41 men) participated in this study for partial fulfilment of course requirements. The mean age was 19.91 years (SD: 1.34). They were selected from different courses (Psychology, Sociology, Pedagogy, Communication and Media, Tourism) from the University of Pécs and Kodolányi János University College in Székesfehérvár and in Budapest.

The North-American sample was recruited from introductory psychology classes at the University of Florida. The number of subjects and their mean age was not provided by the authors (Lang, Bradley & Cuthbert, 2005).

Stimuli

The collection of IAPS pictures has been growing and updated continually. The most current version contains over 900 stimuli. In the present study 239 IAPS color photographs were selected and divided into sets. Each variable (valence, arousal, dominance/control) has three levels: low ($x < 4$), medium/average ($4 \leq x \leq 6$), high ($x > 6$). In this manner, each picture can be represented in a three-dimensional coordinate space. Sets were formed to represent most of the possible combinations of three variables (e.g., negative/low valence, high arousal, high dominance; positive/high valence, medium arousal, medium dominance). A wide variety of semantic categories were presented in each set (e.g., food, social situations, animals, landscape etc.).

Procedure

PowerPoint 2003 was used on a personal computer for presenting the stimuli. A projector was used for visualization. Subjects seated approximately 2-4 m from the screen (approx. 1,5 x 2 m).

The normative rating procedure published in the manual (Lang et al., 2005) was followed. Three pictures served as examples at the beginning of each run. Subjects could practice the rating procedure and determine their own standards for what is a positive, negative and neutral picture. The example pictures were: 2030 (a smiling woman), 7010 (basket) and 3100 (a burn victim). Each picture was displayed for five seconds. Subjects made their ratings immediately after the stimulus presentation. Rating time lasted 15 seconds. During this standard time interval subjects gave ratings on the dimension of valence, arousal and dominance using the Self-Assessment Manikin (SAM) rating system (Lang et al., 2005). The session lasted about one hour.

Results

Ratings of valence, arousal and dominance

The overall mean ratings were 5.32 (SD: 2.00) for valence, 4.72 (SD: 1.41) for arousal and 6.43 (SD: 1.26) for dominance. The following mean ratings were found in the US based sample on approximately 100 subjects: 5.37 (SD: 1.81) for valence, 4.75 (SD: 1.17) for arousal and 5.31 (SD: 1.16) for dominance. Paired t-tests showed no significant difference for valence, $t(238) = -1.29$, $p > .05$) and arousal, $t(238) = 0.67$, $p > .05$). The mean Hungarian dominance ratings were significantly higher than the American standards, $t(238) = 18.17$, $p < .01$) (see Figure 2 and Table 1).

Table 1. Means and standard deviations of ratings³

	All subjects		Men		Women	
	HU	US	HU	US	HU	US
Valence	5.32 (2.00)	5.37 (1.81)	*5.29 (1.85)	5.44 (1.56)	5.33 (2.08)	5.32 (2.04)
Arousal	4.72 (1.41)	4.75 (1.17)	*4.47 (1.54)	4.62 (1.21)	4.79 (1.42)	4.83 (1.20)
Dominance	*6.43 (1.26)	5.31 (1.16)	*6.92 (1.45)	5.53 (1.10)	*6.28 (1.25)	5.13 (1.26)

³ * $p \leq .05$.

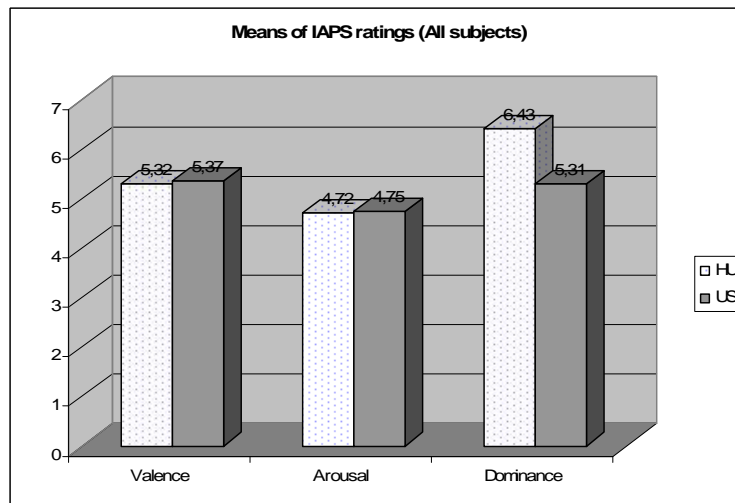


Figure 2. No significant difference was found between Hungarian and U.S. ratings on valence and arousal dimensions. Hungarian subjects showed significantly higher ratings on dominance dimension.

The correlations between valence and arousal for all subjects, male and female, were negative and higher than in the U.S. sample. (For the Hungarian sample: $r = -.54$ /all/; $-.43$ /male/; $-.54$ /female/; $p < .01$. For the U.S. sample: $r = -.26$; $-.04$ (n.s.); $-.35$; $p < .01$).

A positive and lower correlation was found between valence and dominance compared to the U.S. ratings. (For Hungarian sample: $r = .74$ /all/; $.66$ /male/; $.73$ /female/; $p < .01$. For U.S. sample: $r = .81$; $.71$; $.85$; $p < .01$)

We found a negative and stronger correlation between arousal and dominance than in the U.S. sample. (For the Hungarian sample: $r = -.77$ /all/; $-.69$ /male/; $-.77$ /female/; $p < .01$. For the U.S. sample: $r = -.61$ all; $-.52$ male; $-.60$ female; $p < .01$)

We would suggest that the tendencies in the Hungarian and North-American samples are similar. The stronger correlations may be due to the relatively large number of subjects in the Hungarian sample.

The lower the rating on the valence dimension, the higher the score on the arousal scale for the same picture, that is the unpleasant pictures were rated more arousing. A positive correlation between valence and arousal shows that the pleasant pictures correlate with a higher level of control over an affective state, while emotions evoked by unpleasant pictures are more overwhelming. These results are consistent with negative correlations between arousal and dominance. The higher the score on the arousal scale, the lower the rating on dominance.

Affective space

Results of the present study have replicated Lang and colleagues' boomerang-shaped distribution. Figure 3 shows the ratings of 239 IAPS pictures in the affective space determined by valence and arousal dimensions from all subjects in the Hungarian and American sample.

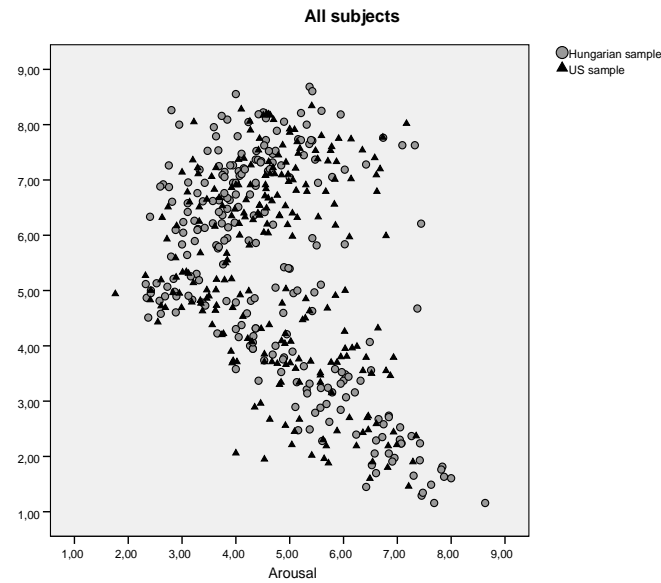


Figure 3. Mean ratings of 239 IAPS pictures for all Hungarian and U.S. subjects. Affective space has been determined by dimensions of valence and arousal. Linear correlation between valence and arousal dimensions for the Hungarian sample: $r = -.54$, $p < .01$, and for the U.S. sample: $r = -.26$, $p < .01$

Figure 4. and 5. represent the distribution of ratings given by Hungarian and U.S. male and female subjects, respectively.

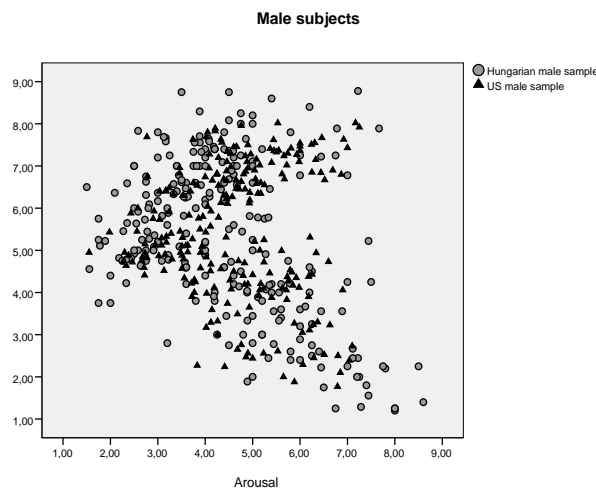


Figure 4. Mean ratings of 239 IAPS pictures for Hungarian and U.S. male subjects. Affective space has been determined by dimensions of valence and arousal. Linear correlation between valence and arousal dimensions for the Hungarian sample: $r = -.43$, $p < .01$, and for the U.S. sample: $r = -.04$, $p > .05$.

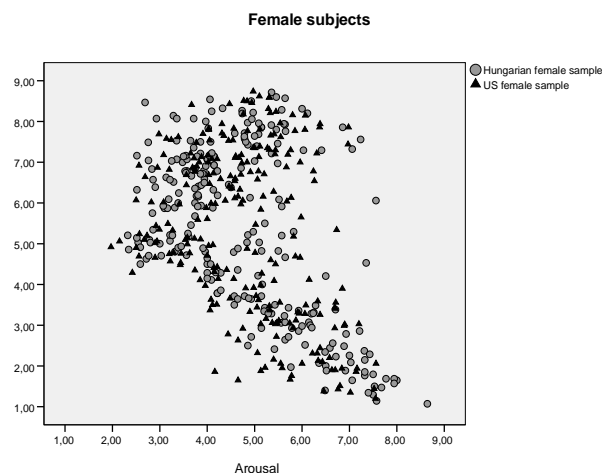


Figure 5. Mean ratings of 239 IAPS pictures for Hungarian and U.S. female subjects. Affective space has been determined by dimensions of valence and arousal. Linear correlation between valence and arousal dimensions for the Hungarian sample: $r = -.54$, $p < .01$, and for the U.S. sample: $r = -.34$, $p < .01$.

Plots form a boomerang-shape in the original American study - that is, pictures that are rated as either highly unpleasant or highly pleasant are also rated as more arousing.

Stimuli were separated into two groups for further examination of motivational systems - approach and avoidance. Sixteen pictures, as neutral stimuli, were excluded on the basis of their valence (between 4.85 and 5.15). The unpleasant picture set consisted of 94 items and the pleasant set contained 129 items. In the Hungarian sample (see Figure 6), a significant positive linear relationship was found between valence and arousal for pleasant pictures ($r = .34$, $p < .01$) and a negative correlation for unpleasant pictures ($r = -.85$, $p < .01$).

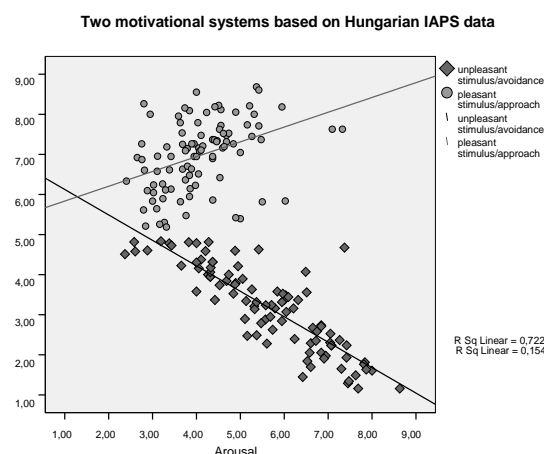


Figure 6. IAPS pictures were separated into pleasant and unpleasant sets. A positive linear relationship was found between valence and arousal for pleasant pictures ($r = .34$, $p < .01$) and a negative correlation for unpleasant pictures ($r = -.85$, $p < .01$).

We conclude that affective states evoked by the separated picture sets (pleasant vs. unpleasant) differ in their pleasantness and arousability. Distribution of the ratings can imply that motivational systems of approach and avoidance have been identified.

The distribution of pictures in the affective space indicates that pictures are suitable for inducing different emotional reactions in participants. Hungarian mean ratings for valence ranged from 1.16 to 8.68, from 2.33 to 8.63 for arousal, and from 2.97 to 8.68 for dominance. North American ratings have varied between 1.46 and 8.34 on the valence dimension. A minimum value of 1.76 and maximum value of 7.35 has been found for arousal. Dominance ratings ranged from 2.15 to 7.71.

Sex differences

Regarding sex differences in the Hungarian sample, no significant difference was found for valence, $t(238) = -0.77$, $p > .05$. Hungarian women's mean ratings for arousal were significantly higher than Hungarian men's scores, $t(238) = -6.15$, $p < .01$. This means that women found the pictures to be more arousing. One possible explanation can be their lower level of threshold for arousal. In regards to the dominance dimension, Hungarian men gave significantly higher ratings compared to Hungarian women, $t(238) = 12.28$, $p < .01$ (see Figure 7.).

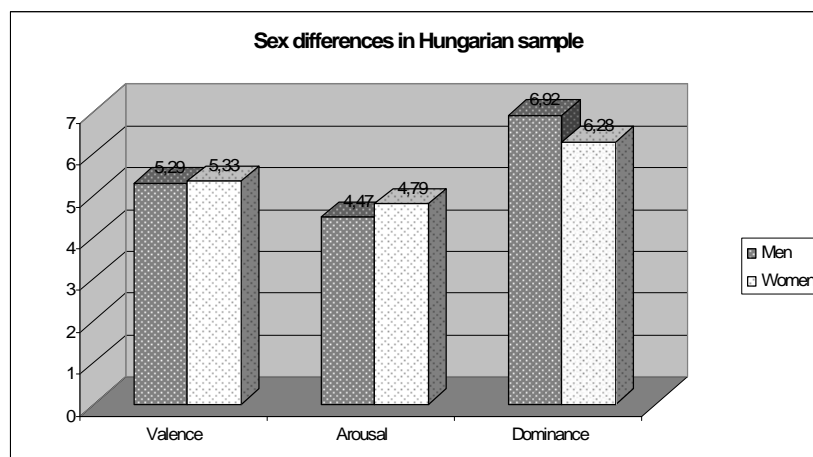


Figure 7. The mean valence ratings of Hungarian men and women did not differ; $t(238) = -0.77$, $p > .05$. Sex differences were found on arousal and dominance dimensions. Hungarian women gave higher arousal scores than men; $t(238) = -6.15$, $p < .01$. Dominance ratings were significantly higher for Hungarian men than women; $t(238) = 12.28$, $p < .01$.

Discussion

A standardized set of stimuli is preferable for cross-cultural studies on emotions. The latest version of the International Affective Picture System (IAPS) (Lang et al., 2005) contains more than 900 color-photos representing a wide variety of semantic contents. The advantage of IAPS is to allow for better control over experimental stimulus selection, thus making experimental results comparable within and across laboratories.

In this study, we compared the Hungarian affective ratings of 239 IAPS pictures with the North-American data. 187 Hungarian university students took part in the experiment. Ratings were given based on valence, arousal and dominance dimensions, using the Self-Assessment Manikin.

Ratings from the Hungarian sample correlated with the U.S. ratings. The highest correlation was found for valence ($r = .95$), followed by arousal ($r = .72$) and dominance dimensions ($r = .69$).

The mean ratings of valence and arousal did not differ between the two groups. The mean ratings of dominance were found to be significantly higher within the Hungarian sample. It seems there have been fast universal appraisal mechanisms evoked by emotional stimuli regardless of the cultural context (Hungarian versus U.S.). In our view, valence scales may represent how subjects assess different stimuli. However, “unpleasantness versus pleasantness” is not the only determinant for emotions. Actual action tendencies should be changed when the level of arousal reaches its threshold. “Approach” and “avoidance” of the stimulus are the two main action tendencies which are connected to the main motivational systems (Lang, Bradley & Cuthbert, 1990; Gray, 1994).

The distribution of ratings in the affective space determined by valence and arousal has underlined the biphasic structure of emotions (Bradley & Lang, 2007b). The upper half of the ratings has supported the assumption of an “appetitive” motivational system, while the bottom half has confirmed the presence of an “avoidance/defensive” motivational system (see Figure 6).

Reactions to emotionally evocative stimuli seem to be universal due to biologically determined evolutionary adaptive mechanisms. However, experiences through learning can have a significant impact on one’s emotional life. Cultural differences should also be taken into account. Since the nature of emotions is determined by both biological and social components, it is a question of whether the universal or the cultural aspect is represented more strongly in our study with IAPS. It has been concluded that biologically determined appraisal mechanisms in assessing the valence and arousal of a stimulus might be universal across cultures, based on our finding of similar ratings on valence and arousal scales between the Hungarian and the U.S. sample. Moreover, the impact of culture and social learning is evident in the significant differences on the dominance dimension. One plausible explanation is that the cultural and social context between Hungary and the United States are vastly different. To test this, samples with similar cultural and historical backgrounds should be used from Central- and Eastern-Europe. Such a study is in progress involving university students from Poland, is currently in progress.

An alternative methodological explanation is the limitation of using the Self-Assessment Manikin technique. In the scientific literature on emotional research, the first two dimensions are discussed more often than the third, though clear reason has been given for this neglect. The dominance dimension and its possible relations with other psychological constructs (e.g., affective style, individual differences in behavioural inhibition and approach system, coping mechanisms) need further examination.

Regarding sex differences, we would like to focus on the results from the Hungarian sample because the exact number of subjects and the male/female ratio in the U.S. sample has not been available. Hungarian men and women did not differ in their ratings of valence. Significant differences were found in their ratings of arousal and dominance. Hungarian women gave higher arousal scores and lower dominance scores than Hungarian men, which may be explained by the different levels of sensitivity of each gender to emotionally evocative stimuli.

In spite of high correlations between the Hungarian and American data, we suggest the usage of Hungarian standards for studies among the Hungarian population. The authors’ future goal is to update the database regularly and give standards for additional IAPS pictures.

References

- Backs, R. W., da Silva, S. P., & Han, K. (2005). A comparison of younger and older adults' Self-Assessment Manikin ratings of affective pictures. *Experimental Aging Research*, 31, 421-440.
- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'Reading the Mind in the Eyes' Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42, 241-251.
- Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a "language of the eyes"? Evidence from normal adults and adults with autism or Asperger syndrome. *Visual Cognition*, 4, 311-331.
- Bensafi, M., Rouby, C., Farget, V., Bertrand, B., Vigouroux, M., & Holley, A. (2002). Autonomic nervous system responses to odours: The role of pleasantness and arousal. *Chemical Senses*, 27, 703-709.
- Bensafi, M., Rouby, C., Farget, V., Bertrand, B., Vigouroux, M., & Holley, A. (2003). Perceptual, affective, and cognitive judgments of odours: Pleasantness and handedness effects. *Brain and Cognition*, 51, 270-275.
- Bensafi, M., Rouby, C., Farget, V., Vigouroux, M., & Holley, A. (2002). Asymmetry of pleasant vs. unpleasant odour processing during affective judgment in humans. *Neuroscience Letters*, 328, 309-313.
- Bradley, M. M. & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavioral Therapy & Experimental Psychiatry*, 25, 49-59.
- Bradley, M. M. & Lang, P. J. (2007a). *The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. Technical report B-3*. Gainesville, FL: University of Florida.
- Bradley, M. M., & Lang, P. J. (2007b). Motivation and emotion. In J. T. Cacioppo, L. G. Tassinari, & G. Berntson (Eds.), *Handbook of psychophysiology* (pp. 581-607). New York: Cambridge University Press.
- Bradley, M. M. & Lang, P. J. (2007c). The International Affective Picture System (IAPS) in the study of emotion and attention. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment*. Oxford University Press.
- Breiter, H. C., Etcoff, N. L., Whalen, P. J., Kennedy, W. A., Rauch, S. L., Buckner, R. L., Strauss, M. M., Hyman, S. E. & Rosen, B. R. (1996). Response and habituation of the human amygdala during visual processing of facial expression. *Neuron*, 17, 875-887.
- Clark, L. A. & Watson, D. (1994). Distinguishing functional from dysfunctional affective responses. In P. Ekman & R. J. Davidson (Eds.), *The Nature of Emotions* (pp. 131-136). New York: Oxford University Press.
- Damasio, A. R. (2000). A Second chance for emotion. In R. D. Lane & L. Nadel (Eds.), *Cognitive Neuroscience of Emotions* (pp.13-23). Oxford University Press.
- Darwin, Ch. R. (1872/1963). *The expression of the emotions in man and animals*. London: John Murray.
- Ekman, P., Friesen, W. V. & Hager, J. C. (2002). *Facial Action Coding System. The Manual on CD ROM*. Research Nexus division of Network Information Research Corporation, USA.
- Fitzgerald, D. A., Angstadt, M., Jelsone, L. M., Nathan, P. J., & Phan, K. L. (2006). Beyond threat: Amygdala reactivity across multiple expressions of facial affect. *NeuroImage*, 30, 1441-1448.

- Frijda, N. H. (1994). Emotions are functional, most of the time. In P. Ekman & R. J. Davidson (Eds.), *The Nature of Emotions* (pp.112-122). New York: Oxford University Press.
- Gray, J. A. (1994). Three fundamental emotion systems. In P. Ekman & R. J. Davidson (Eds.), *The Nature of Emotions* (pp.243-247). New York: Oxford University Press.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review*, 97, 377-398.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-6*. Gainesville, FL: University of Florida.
- Levenson, R. W. (1994). Human Emotion: A Functional View. In P. Ekman & R. J. Davidson (Eds.), *The Nature of Emotions* (pp. 123-126). New York: Oxford University Press.
- Levenson, R. W., Soto, J., & Pole, N. (2007). Emotion, biology and culture. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp.780-796). New York, London: The Guilford Press.
- Martinez, A. M. & Benavente, R. (1998). *The AR Face Database. CVC Technical Report #24*, June.
- Mesquita, B. & Leu, J. (2007) The Cultural psychology of emotion. In S. Kitayama & D. Cohen (Eds.), *Handbook of cultural psychology* (pp.734-759). New York, London: The Guilford Press.
- Morris, J.S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, 393, 467-470.
- Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, 1, 261-289.
- Osgood, C. E. (1962). Studies on the generality of affective meaning systems. *American Psychologist*, 17, 10-28.
- Öhman, A., Flykt, A., & Lundqvist, D. (2000). Unconscious emotion: Evolutionary perspectives, psychophysiological data, and neuropsychological mechanisms. In R. D. Lane & L. Nadel (Eds.), *Cognitive Neuroscience of Emotions* (pp. 297-327). Oxford University Press.
- Panksepp, J. (2006). Emotional endophenotypes in evolutionary psychiatry. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 30, 774-784.
- Plutchik, R. (1994). The language of emotions. In *The Psychology and Biology of Emotions* (pp. 45-72), Harper Collins College Publishers.
- Ribeiro, R. L., Pompeia, S., & Bueno, O. F. A. (2005). Comparison of Brazilian and American norms of International Affective Picture System (IAPS). *Revista Brasileira de Psiquiatria*, 27, 208-215.
- Scherer, K. R. (1994). Emotion serves to decouple stimulus and response. In P. Ekman & R. J. Davidson (Eds.), *The Nature of Emotions* (pp. 127-130). New York: Oxford University Press.
- Suslow, T., Ohrmann, P., Bauer, J., Rauch, A.V., Schwindt, W., Arolt, V., Heindel, W., & Kugel, H. (2006). Amygdala activation during masked presentation of emotional faces predicts conscious detection of threat-related faces. *Brain and Cognition*, 61, 243-248.
- Verschuere, B., Crombez, G., & Koster, E. (2001). The international affective picture system: A Flemish validation study. *Psychologica belgica*, 41, 205-217.
- Wundt, W. (1896/1897). *Grundriss der psychologie. (Outlines of Psychology)*. Leipzig: Engelmann. (<http://psychclassics.asu.edu/Wundt/Outlines>)

Life expectancy and psychological immune competence in different cultures

Attila Oláh¹, Henriett Nagy, Kinga G. Tóth

Eötvös Lorand University, Budapest, Faculty of Education and Psychology
Department of Personality and Health Psychology
46. Izabella str., Budapest, Hungary H-1064

Abstract

In our study we examined the relationship between psychological immune competence and life expectancy at birth, plus the relationship between the life expectancy at birth and per-capita GDP. We supposed that a health promoting culture and material investment in health promotion will influence the level of psychological immune competence and thus contribute to an increase of life extension. Our sample consisted of more than three thousand young participants from 12 different countries. We administrated the Psychological Immune Competence Inventory (Oláh, 2005) to measure psychological immune competence and used the World Factbook 2006 to determine life expectancy and per-capita GDP. We found that there is a strong correlation between psychological immune competence and life expectancy, and this relationship is stronger than the correlation between life expectancy and per-capita GDP. As our conclusion states, these results prove that culture which implants psychological immunity contributes to life extension even more than material investment.

Keywords: Life expectancy, psychological immune competence, per-capita GDP

Introduction

The mainstream of life expectancy research focuses on the role of socio-demographic factors. These studies demonstrated that in developed countries women live longer than men, and higher socio-economic status is associated with better health conditions and longer lifetime (see Sarafino, 1990; Kennelly, O'Shea & Garvey, 2003; Mathers et al, 2004). However, up till now relatively scarce evidence is available regarding the psychological mediating factors of life expectancy. In 1993 Friedman, Tucker, Tomlinson-Keasey, Schwartz, Wingard and Criqui found a strong positive correlation between life expectancy and conscientiousness, which is defined as the propensity to follow socially prescribed norms and rules regarding impulse control, to be goal directed, planful, and be able to delay gratification (John and Srivastava, 1999).

¹ E-mail: olah.attila@ppk.elte.hu

Pressman and Cohen (2005) called attention to the important mediatory role of positive affect between personality traits and health. Maier and Smith (1999) found that lower well-being entails a higher mortality risk. In an extensive study Veenhoven (2005) compared the quality of life in 67 different countries. He measured how long people live and how happy they are. Combining these measures he formed a Happy-Life-Years indicator, which showed significant differences between the examined nations. According to his results, economic situation, freedom and justice are able to account 66% of the variance of happy life years.

To date, reviews examining associations between positive psychological constructs and health outcomes have been broad in scope, including traits such as self-esteem, extraversion, purpose, mastery, and optimism (Lyubomirsky, King, & Diener, 2005; Ryff, 2003; Salovey, Rothman, Detweiler, & Steward, 2000). Several recent studies demonstrated that the protective traits of personality determine healthy behaviour, and the developmental level of protective traits is influenced by cultural factors (e.g., sense of coherence (Antonovsky, 1987) positive thinking (Lee & Seligman, 1997) self-control (Peele, 1989) self-efficacy (Scholz, Dona, Sud & Schwarzer, 2002)).

In our study we examine the relationship between life expectancy estimated at birth and psychological immune competencies (such as positive thinking, sense of control, sense of coherence, change and challenge orientation, sense of self-growth, social monitoring capacity, creative self-concept, self-efficacy, goal orientation, synchronicity, impulse control, emotional control and irritability control), plus the relationship between life expectancy at birth and per-capita GDP.

Psychological immune system

The psychological immune system (PIS, Oláh, 2005) is a promotive and preventive system, which integrates psychological immune competencies. It optimizes and strengthens those processes of the interaction between the individual and his/her environment which primarily serve the self. The common characteristics of the components of the psychological immune system are the following: (1) they tune the cognitive apparatus toward the perception of possible positive outcomes, (2) they strengthen the anticipation of the possible success of behavior, (3) they contribute to positive changes in one's state and emphasize developmental opportunities, (4) they guarantee the selection of coping strategies which fit both the characteristics of the situation and one's own state and disposition, and finally (5) they ensure the monitoring of the individual's coping resources and their quick and adequate mobilization.

Specifically, the PIS functions as a superordinate system with three interacting subsystems. Some components of the psychological immune system - such as positive thinking, sense of control, sense of coherence, change and challenge orientation and sense of self-growth - directly influence primary appraisal. These accommodative competencies create the first subsystem of PIS, which guides the organism's orientation towards the environment, and facilitates the appraisal of the environment in a meaningful way, as a positive and manageable entity. The assimilative components of PIS - such as social monitoring capacity, creative self-concept, self-efficacy and goal orientation - directly influence the process of secondary appraisal. This set creates the second subsystem of PIS, which initiates the seeking out and the assimilation of information, and subsequently it puts the necessary resources into action in order to influence the environment and to create new possibilities. Finally, the regulating components of PIS - such as synchronicity, emotional control, and irritability control - create the third subsystem, which insures the functioning of the first two subsystems by stabilizing the individual's inner emotional life.

The psychological immune system strengthens the position of the self from the beginning of the coping process, controlling it from the first step (i.e. the cognitive evaluation of the threats through the selection and activation of behavioural responses, and terminates in the self's elevated resistance to threats and stress). It is a system that acts similarly to the biological immune system, mediating the impact of psychological stress. This system not merely controls the coping processes, but it is also responsible for the effectiveness, integrity and growth potential of the person, gathering and synchronizing the resources of personality: traits, knowledge and experience connected to active dealing with stress (Oláh, 2005).

Oláh (2005) compared the PIS profiles of cancer patients, coronary patients and alcohol-addicts with the PIS profiles of a sample of normal individuals. The data revealed both physical and psychic pathological status is in connection with the weaknesses of the PIS subsystems and corresponding constitutional factors, whereas a strong and integrated PIS correlates with the frequency of flow experiences, a higher level of well-being and life-satisfaction, and with symptom-free status.

Research question

In the present study, we hypothesized that health promoting culture and material investment in health promotion will influence the psychological immune competence of the youth growing up in the given country, and these two factors will contribute to an increase of their life expectancy. We hypothesized that immune competence correlates with life expectancy estimates at birth, and in accordance with previous research data, we hypothesized, that there would be a strong correlation between life expectancy at birth and per-capita GDP.

Methods

Participants

Altogether 3878 persons (mean age = 21.58, SD = 1.43) took part in our research, 1909 males (mean age = 21.3, SD = 1.23) and 1969 females (mean age = 21.8, SD = 1.34). Three hundred twenty-four persons live in Russia (159 males, mean age = 21.4, and 165 females, mean age = 21.8), 308 in Sweden (152 males, mean age = 22.02 and 156 females, mean age = 22.2), 306 in Norway (148 males, mean age = 22.3 and 158 females, mean age = 22.6), 338 in Yemen (174 males, mean age = 20.9 and 164 females, mean age = 21.2), 330 in Germany (160 males, mean age = 22.7, and 170 females, mean age = 22.3), 363 in India (180 males, mean age = 21.7, and 183 females, mean age = 21.9), 377 in Hungary (185 males, mean age = 22.1, and 192 females, mean age = 22.3), 373 in Romania (182 males, mean age = 21.6, and 191 females, mean age = 21.9), 293 in Cyprus (142 males, mean age = 21.9, and 151 females, mean age = 22.2), 270 in Israel (137 males, mean age = 22.6, and 133 females, mean age = 22.8), 271 in Mauritius (129 males, mean age = 21.8, and 142 females, mean age = 22.1), and 325 in Greece (161 males, mean age = 22.2, and 164 females, mean age = 21.9).

Measures

We administered the Psychological Immune Competence Inventory (PICI, Oláh, 2005) to measure positive personality traits. PICI consists of 80 items that make up 16 dimensions and 3 factors. All of these dimensions consists of 5 items, some of them reversed. All items are a

4 step likert scale, ranging from 1 to 4. The PICI was translated to different languages, and all of the participants completed the test using their native language (see Figure 1).

1	Οι άνθρωποι με περιγράφουν σαν πολύ αισιόδοξο άτομο				
2	Σύμφωνα με την δικιά μου εμπειρία, η επιτυχία είναι αποτέλεσμα καλού σχεδιασμού.				
3	Όταν κοιτάω πίσω στο παρελθόν μου και μπροστά στο μέλλον μου, κατανοώ την ζωή μου ικανοποιητικά μέσα από τις εμπειρίες της.				
4	3 2 1	1	2	3	4
4	3 2 1	1	2	3	4
4	3 2 1	1	2	3	4
4	3 2 1	1	2	3	4
1.	Люди считают меня оптимистом	1	2	3	4
2.	Я считаю, что удача – это результат хорошего планирования.	1	2	3	4
3.	Я понимаю сущность своей жизни.	1	2	3	4
4.	Я очень довольна (доволен) собой и тем, чего Я уже достигла (достиг) в жизни	1	2	3	4

Figure 1: Items from PICI in different languages

- (1: People describe me as a very optimistic person,
 2: According to my experience, success is a result of good planning,
 3: When I look to my past and to my future, I view that my life is understandable,
 4: I am very happy with myself and what I have accomplished in life)

Procedures

Participants completed the *PICI* in groups using their native languages. Besides employing *PICI* we also collected information about participant's age and gender. We analyzed these data with using the SPSS 16.0 statistical package. We computed correlations between the immune competence total scores, life expectancy at birth, and per-capita GDP (to determine the per-capita GDP we used the World Factbook, 2006, CIA).

Results

First, reliability tests were run, and it was found that Cronbach-alphas of the *PICI* scales were satisfactory, having a range from 0.84 to 0.88. The descriptive statistics of psychological immune score, life expectancy at birth and GDP-per-capita in twelve different countries are summarized in Figure 2, Figure 3 and Table 1. Bivariate correlations were conducted to test the relationship between psychological immune competence and life expectancy at birth, plus between psychological immune competence and per-capita GDP. The correlational analysis revealed a strong positive relation between the psychological immune competence scores and life expectancy at birth (among males: $r = 0.68$, $p < 0.01$; among females: $r = 0.71$, $p < 0.05$), and a moderate positive relation between psychological immune competence and per capita GDP (among males: $r = 0.47$, $p < 0.05$, and among females: $r = 0.42$, $p < 0.05$).

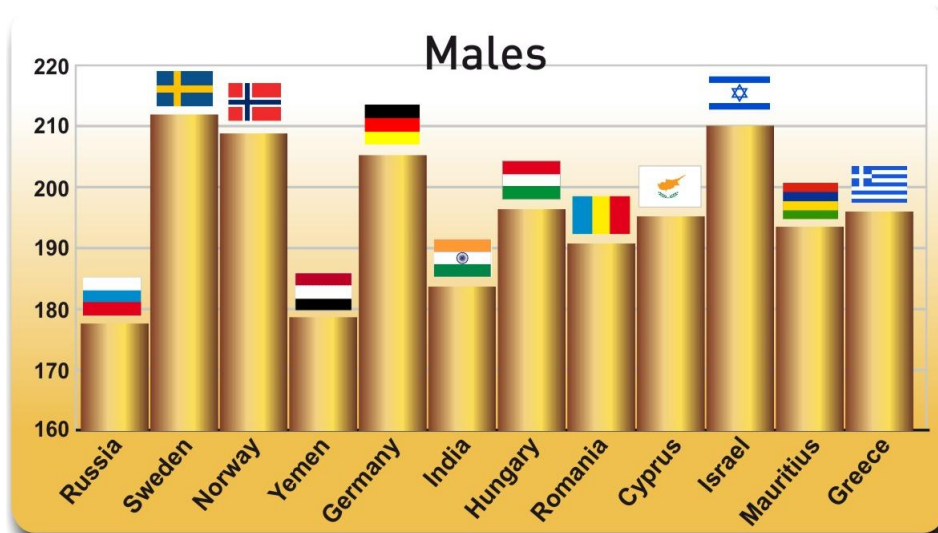


Figure 2: Levels of psychological immunity among males in different countries

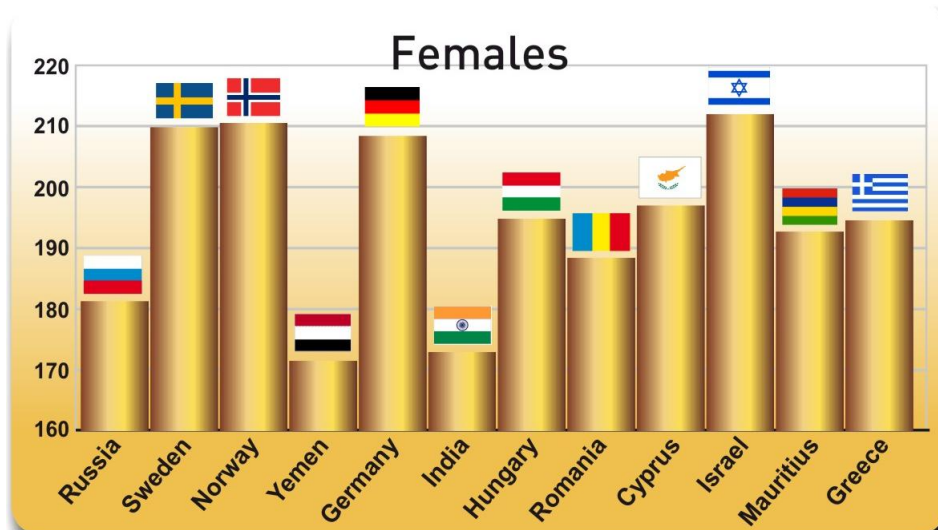


Figure 3: Levels of psychological immunity among females in different countries

Table 1: Psychological immune scores, life expectancy at birth and per-capita GDP in different cultures

Country	Life expectancy at birth* Males	Life expectancy at birth* Females	GDP/capita	Psychological Immunity Males	Psychological Immunity Females
Russia	59.91	73.27	8.900\$	177.4	181.4
Sweden	78.12	82.62	26.800\$	211.6	209.7
Norway	76.74	82.01	37.800\$	209.2	210.5
Yemen	59.53	63.29	800\$	178.5	171.6
Germany	75.56	81.68	27.600\$	204.4	208.4
India	63.25	64.77	2.900\$	183.5	172.5
Hungary	68.07	76.69	13.900\$	196.5	196.4

Romania	67.63	74.82	7.000\$	190.4	188.5
Cyprus	75.29	80.13	20.300\$	195.3	196.8
Izrael	77.08	81.37	19.800\$	210.3	211.3
Mauritius	68.11	76.13	11.400\$	194.3	192.5
Greece	76.59	81.76	21.300\$	196.4	195.4

*Reference: The World Factbook, 2006, CIA

Table 2: The correlations between psychological immune competence, and life expectancy at birth and per- capita GDP

	Life expectancy at birth	
	Males	Females
Psychological immune competence	0.68**	0.71**
GDP/capita	0.47*	0.42*

1909 Males; 1969 Females; * = $p < 0.05$; ** = $p < 0.01$

Discussion

The focus of the present study was to explore the relationship between psychological immune competence and life expectancy, and in accordance with other studies on life expectancy and socio-demographic features, we hypothesized that life expectancy would correlate with per-capita GDP. The correlations of positive personality traits and life expectancy were examined among 1909 males and 1969 females from 12 different countries.

The significant, strong positive relationship between life expectancy and psychological immunity confirmed our first hypothesis. In accordance with other studies we found that per-capita GDP was also significantly related to life expectancy. Although per-capita GDP was also significantly related to life expectancy, these correlations were not as strong as the correlations of psychological immunity and life expectancy. Thus, our results support the notion that culture, which fosters psychological immunity, contributes to life extension even more than material investment. According to our results the cultural factors seem to be more important than socio-demographic factors in the area of life expectancy.

The limitations of the study should be considered. Data were collected through the use of a single survey at a single point in time. Longitudinal research designs would be necessary in order to understand more precisely the relationship between life expectancy and per-capita GDP. The second important limitation is that the sample size only permitted correlational analyses in our study, which do not provide causal explanations for the findings. Certainly, these findings should encourage further research, focusing on cultural differences in life expectancy. The implications of the findings for practice may be expanded with future research.

References

- Antonovsky, A. (1987). *Health, Stress, and Coping*. San Francisco: Jossey-Bass.
- Friedman, H. S., Tucker, J. S., Tomlinson-Keasey, C., Shwartz, J. E., Wingard, D. L., & Criqui, M. H. (1993). Does childhood personality predict longevity? *Journal of*

- Personality and Social Psychology*, 65, 176-185.
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (Vol 2, pp.102-138). New York: Guilford Press.
- Kennelly, B., O'Shea, E., & Garvey, E. (2003). Social capital, life expectancy and mortality: a cross-national examination. *Social Science & Medicine*, 56(12), 2367-2377.
- Lee, Y., & Seligman, M. E. P. (1997). Are American more optimistic than Chinese? *Personality and Social Psychology Bulletin*, 23, 32-40.
- Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success? *Psychological Bulletin*, 131, 803-855.
- Maier, H., & Smith, J. (1999). Psychological predictors of mortality in old age. *Journal of Gerontology: Psychological Sciences*, 54, 44-54.
- Mathers, C. D., Iburg, K. M., Solomon, J. A., Tandon, A., Chatterji, S., Ustin, B., & Murray, C. J. L. (2004). Global patterns of healthy life expectancy in the year 2002. *BMC Public Health*, 4, 66.
- Oláh, A. (2005): *Anxiety, coping, and flow. Empirical studies in interactional perspective*. Budapest: Treffort Press.
- Peele, S. (1989). *The diseasing of America*. Boston: Houghton Mifflin.
- Pressman, S. D., Cohen, S. (2005). Does positive affect influence health? *Psychological Bulletin*, 131(6), 925-971.
- Ryff, C. D. (2003). The role of emotion on pathways to positive health. In R. Davidson (Ed.), *Handbook of affective sciences* (pp. 1083-1104). London: Oxford University Press.
- Salovey, P., Rothman, A. J., Detweiler, J. B. & Steward, W. T. (2000). Emotional states and physical health. *American Psychologist*, 55, 110-121.
- Sarafino, E. P. (1990). *Health psychology. Biopsychosocial interactions*. Wiley, New York.
- Scholz, U., Dona, B. G., Sud, S., Schwarzer, R. (2002). Is general self-efficacy an univesal construct? *European Journal of Psychological Assessment*, 18, 242-251.
- The World Factbook (2006). <https://www.cia.gov/library/publications/the-world-factbook/>
- Veenhoven, R. (2005). Apparent quality of life in nations: How long and happy people live. *Social Indicators Research*, 71(1), 61-86.

Hungarian answers to an international crises: Parliamentary debates during the 1999 Kosovar conflict

Paszkal Kiss*

Eötvös Lorand University, Budapest, Faculty of Education and Psychology
Department of Social and Educational Psychology
46. Izabella str., Budapest, Hungary H-1064

Abstract

This article examines different contexts of political discourse in Hungary in a heated situation of Hungarian participation in the NATO interventions to the Kosovar conflict. The Hungarian position had many complexities (Hungarian minority in Serbia, neighbouring relations, aspirations to join the Transatlantic community) that were reflected in the dynamics of political discourse. Overall representations of international relations, their historic change, and the national perspective (aspirations) were seen to determine the views of political actors and ultimately their behaviour. It was also important to show that even in such a heated situation the complexity of intergroup relations appearing in the discourse of the opposing sides was not seen as black and white and the shades changed dynamically through time.

Keywords: foreign policy, social representations, Hungary, political discourse, Kosovar conflict

Introductory remarks

As cross-fertilization between psychology and political science, social psychological concepts and theories are used in understanding foreign policy decisions (Levy, 2003) with growing success. Representations of the nation, EU and Europe will be analysed in this paper against the background of more inclusive international categories (the Euro-Atlantic region, a globalised world) in an attempt to further this line of thought. Identification will be taken as a process of self- and ingroup-categorization into social groups on different levels of abstraction, e.g. nations, Central Europe, European Union, Europe (Kiss, 2002, 2005).

We focus here on foreign policy issues as they are assumed to reveal the relation between nation, EU and Europe. Before Hungary's 2004 accession to the European Union there was a generally accepted set of three priorities or principles in Hungarian foreign policy established during the democratic changes in 1989-1990: (1) Euro-Atlantic integration; (2) Regional politics (good neighbourly relations); (3) protection of Hungarian minorities in the Carpathian Basin (integrating the Hungarian nation across borders). These principles have

* kiss.paszkal@ppk.elte.hu

set a consensual framework for Hungarian foreign policy, but they also give ground for differences between political parties and governments as they are in conflict at times.

This qualitative study within a broader research program (Euronat research project¹) was conducted within the general theoretical framework of Social Representation Theory (Moscovici, 1984). This theory considers the social-societal processes involved in shaping public representations of new social phenomena. Public and even mass communication processes are central to establishing and re-establishing representation, through agreement and difference (Canter & Monteiro, 1993). Even if many times the core of representations is built from general consensus, this itself also provides the dimensions for group (or individual) differences. Doise (Doise et al., 1993) describes this secondary process of differentiation as an anchoring of differences to group membership. These differences will be reconstructed in different perspectives to the Kosovar conflict and the role of Hungary in it.

Consensus and differences can both be seen in parliamentary discussions about foreign policy as international relations often are based on a widely accepted view about national interest and values. The representations of nation and Europe (European Union) in political discourse may be blurred however for several reasons: (1) National identity tends to be implicit in debates. Nation itself is rarely thematised, often appearing in relation to other policy questions (e.g., cultural or foreign policy). Given the diversity of these topics, it is necessarily restrictive to pick up some for analysis. (2) We have to be cautious when analysing the appearance of these general (thus vague) and positively evaluated concepts in political debates as they are part of a thematising struggle between political parties. (3) They might also be merely 'rhetoric' in justifying a policy position taken (Reicher, Hopkins & Jost, Condor, 1997), although national and European identification are assumed to affect profoundly the political actions taken. (4) A related issue is that these representations are vague and positive enough to be loaded by different ideological connotations. The same expression may carry different meanings when used by different political parties and can be associated to different values.

Hungarian political parties, parliamentary system

It was with the 1990 political changeover that the institutional bases of the multiparty pluralistic parliamentary democracy were laid down, in the preparation of which the opposition movements growing in strength from the mid-eighties had taken an active part. The communist party leaders who consented to having talks and the negotiating delegation of opposition members together carried out the change of political systems with a 'constitutional revolution'. The legal traditions of Hungarian politics helped to imbue the previous dictatorship's constitution with completely new meaning and in establishing the legal framework of the multi-party political system. This radically transformed the system of political institutions, and while agreeing with the representatives of the dictatorship, it resulted in limited personal changes in non-political areas. In the sphere of economics, however, the political elite of the dictatorship grasped strong positions through the so-called 'spontaneous privatisation'. Therefore, for a long time it was a key question of political life whether there could be competing elites within Hungarian society. After the political changes, parliamentary forces, especially the outstandingly popular president and the newly founded Constitutional Court (now president of the republic), whose activity gained international recognition, played a major role in further building the constitutional system.

¹ <http://www.eui.eu/RSCAS/Research/EURONAT/>

Of the parties of political change, the national-conservative Hungarian Democratic Forum (MDF) and the liberal Alliance of Free Democrats (SZDSZ) gained the largest voting block in the first free elections. In the national-conservative camp, there were the 'historic' parties that undertook to maintain the heritage of parties of the pre-communist era (FKGP – the smallholders' party and KDNP – the Christian Democrats). In the liberal camp, a generational party, namely the Alliance of Young Democrats also got into Parliament. The Hungarian Socialist Party, a new name for the social democratic party that had been organised from the former communist party, had little weight and was slightly excluded from the work of the first Parliament. Over the subsequent twelve years, each four years different parties had the chance to form coalition governments. It was from 2002 that a socialist-liberal government could remain in power for two terms although with serious loss of popular support after their re-election. Hungarian parties have undergone dramatic change and their supporters have changed as well in the last two decades of parliamentary democracy. For some time now, the decisive party of the left has been the Hungarian Socialist Party, while by the 2002 elections the two major parties of the political changeover (MDF and SZDSZ) have lost so many of their supporters that they hardly managed to get into Parliament. Three of the reasons for these changes should be mentioned here.

The parties of the first free elections in 1990 were established within weeks or months. Their social embeddedness was very weak as they had been born out of intellectual sub-cultures. Another reason for the stormy changes of the party system has been the fact that adjusting to the political changeover and the new socio-economic situation has resulted in economic fallback of nearly a decade and consequently in wide-ranging social discontent. The most recent and probably most significant change in the Hungarian party system took place in the year 1996 when FIDESZ, which had been radically liberal before, took the lead at the head of the rather disjointed right and changed its name to 'FIDESZ – Hungarian Citizens' Party'. Their coming into government at the 1998 general elections, has created deep changes in the Hungarian party system, giving rise to the realistic alternative of the unifying right against the left.

Meanwhile, the party system undergoing such enormous transformations did not entail government crises. All the elected governments have filled their 4-year terms and have ensured the stability of administration. However, after the 2006 "lie-speech" of prime minister Gyurcsány serious questions were raised about the legitimacy of the government by the opposition and social 'unrest' including street fights accompanied the falling of popular confidence in the government.² The Hungarian Parliament has one chamber, with political parties representing citizens. At elections, the threshold for getting seats is winning at least 5 per cent of the mandates. The Hungarian election system is rather complicated, as it is a mix of direct and list election. In the first round, votes are cast for parties' lists (a total of 210 mandates) and individual candidates (176 mandates). If no individual representative wins at least half of all valid votes (which is made quite possible by too many parties being involved), then in the second round individual candidates are repeatedly chosen from. The two rounds enable coalition negotiations and parties' co-operation. Another characteristic of the Hungarian parliamentary system is the relatively high number of two-thirds laws, which necessitates agreement between the government and its opposition.

Between 1998 and 2002 a right-wing coalition (FIDESZ, FKGP, MDF) formed the government in Hungary. In opposition were the socialists, the liberals, and an ultra-right party, the *Party of Hungarian Justice and Life* (MIÉP) got into Parliament. Its views often fall beyond the political consensus and its rhetoric is related to that of other extremist European parties.

² For further details see: http://en.wikipedia.org/wiki/2006_protests_in_Hungary

Research design

Sources

Parliamentary debates on a specific policy issue (the Kosovar conflict) were taken as applications of the foreign principles and reflections of national identity and international attitudes of the Hungarian political elite. This source is regarded as discursive not only because of the immediate parliamentary context, but also because it was framed within a situation of contest in the political arena. Furthermore, the relation between arguments and political actions could be tested in the parliament when the discussion leads to accepting policy proposals.

Parliamentary debates were analysed in a written format. Although it is rich in non-verbal communication, here we concentrated on the content of representations, thus we used the written Journals of the Parliament (www.parlament.hu/napló) as our source.

Parties included in the analyses

The necessarily restrictive nature of this analysis does not allow us to include all significant political parties. Therefore the largest parties were chosen from the left and the right (i.e. MSZP, FIDESZ) and two smaller parties that had a characteristic opinion on national and European issues (SZDSZ and MIÉP). These parties probably differ the most on national or international issues. Other parties on the political right have an opinion differing only in shades from that of FIDESZ. On the left, there is only one noteworthy party (not represented in the parliament) besides MSZP and SZDSZ, that is the communist MSZMP. Its views have only a marginal effect on the general public, however.

Analysis dimensions and contexts considered

In the analysis of Hungarian political discourse we paid attention to a handful of dimensions in building representations together with considering the possible constraints of this discourse set by the most relevant contexts. As for dimensions of analysis we concentrated on the interrelatedness of “State”, “Nation”, “EU”, “Europe” and “the globalised world”. A major part of these interconnections is represented in the threefold priorities of Hungarian foreign policy (helping Hungarian minorities in neighbouring countries, good neighbourly relations, Euro-Atlantic integration of Hungary). The dynamics among regional, national and European identification have been studied extensively (see Doise, Devos, 1999 for a summary) showing the interrelatedness of these identity levels. Major changes and crises (EU accession and Kosovar conflict) were used as important ‘boundary’ cases to reveal the underlying structure of representations behind forming political strategy.

The historical is an especially important context for the studied fragment of political discourse. The Hungarian national identity and its relation to other nations is sketched elsewhere (Kiss & Hunyady, 2005). Here I would emphasize the understanding of Europe’s historical structure as unequally developed zones in interaction (Kosáry, 2001). The (Transatlantic) West was represented positively by Hungarian political elites well before the democratic changes. This was one of the factors indeed preparing Hungary’s smooth adaptation when free political action became possible (Csizmadia, 2001). The Balkan conflicts in this historic change were seen as tests of Hungarian preparedness to join this community even by the Hungarian political elite (Granasztói, 2000). Hungarian perception of the international community deserves a bit more detailed elaboration. Generally, social-economic development

of Hungary and its Euro-Atlantic co-operation are seen as intertwined, but there are also critical views about the EU enlargement and NATO membership. Conflicts and armed conflicts especially play an important role as borderline situations (see Szabó, 2000 referring to Schmitt's approach in analysing the political) in international politics. In these situations the differentiation between 'us' and 'them' is accentuated but that focus does not always mean that they are necessarily turned black and white (Kiss, 2001). Another interpretative context of the political discourse between political parties is that of the public opinion, which sometimes has a pronounced effect on political positions (e.g., in election campaigns or in challenging situations). Whether a party is in government or opposition is also an important background or frame to understand political discourse, especially in parliamentary debates. As a general consequence, the opposition is often more critical about necessary steps and governing parties are defending them.

Hungarian response to the Kosovar conflict

The Hungarian Parliament discussed the Hungarian policy on the Kosovar conflict several times. A decree was accepted on 14 October, 1998 that served as a general legal-political framework for the Hungarian participation³. It was modified⁴ on 24 March, 1999 extending the Hungarian involvement. On 12 April, 1999, the prime minister informed the parliament about the current state of the NATO intervention aside agenda, after which a brief discussion followed. In May-June MSZP attempted to narrow down the Hungarian support for NATO strikes, excluding air strikes from Hungarian territory by proposing a modification of the Decree on 1 June, 1999. The proposal was briefly discussed and was not accepted as an item on the agenda⁵. On 14 June, 1999, the foreign minister informed the Parliament about the peace-declaration and a brief discussion followed.

The general discussion of the 1998 Decree was regarded as setting the interpretative framework of discourse. Analyses of subsequent debates were organized around these positions, searching for additional elements of the original perspectives and taking chronological changes into account as an important factor.

The 1998 Decree

All speakers highlighted that decisions made in this conflict are of highest importance for Hungary. The 1998 Decree was even described as a "historic parliamentary decision" (Gyuricza, FIDESZ: 19/17*) showing Hungary's commitment as a future member of NATO (Hungary was accepted as full member on 12 March 1999) and serving Hungary's strategic interest in having long-term peace in the neighbourhood. Its consequences were seen to be affecting "Hungary's actual security, its long term security interests, the Hungarians living in Vajdaság (Voivodina)⁶, and also Hungary's relation to the NATO" (Kovács, MSZP: 21/17) As generally in foreign policy and especially in international conflicts, there was a high degree of agreement between political parties. A peace was generally seen to be possible together with democratic change in Yugoslavia, where human and minority rights are both safeguarded. All sides emphasized that a lasting solution can be reached only through negotiations and mutual

³ „Decree on Hungary's Contribution to NATO's Action to Solve the Kosovar Crisis" (H/1034) accepted by a 259:11 majority (abstained 6).

⁴ The modified Decree was accepted by a 255:12 (abstained 8) majority.

⁵ A majority of 200:84 (abstained 10) put the proposal off agenda.

* No. of contribution/No. of Parliamentary session

⁶ There are still around 300,000 Hungarians in Northern Yugoslavia, in a more or less circumscribed territory.

agreements. The possible NATO intervention was interpreted as a necessary outside force pushing these negotiations forward. Albeit the consensual support for NATO intervention, Hungary was considered to be in a peculiar situation for several reasons. Firstly, Hungary should not get involved in the conflict only if the lives of Hungarians (on both sides of the borders) are sufficiently protected. Secondly, the Hungarian involvement should not reach a point where it triggers a counter-attack on Hungary (or on Hungarian minorities) and endangers good-neighbourly relations with Yugoslavia in the future.

Despite the general agreement, differences were also discernable in interpretations given to the situation and in reasons for supporting common actions. The small fraction of MIÉP was the only party to break consensus by not supporting the Decree. They interpreted the whole situation exclusively within a politics of the great-powers. They believed the Russian encouragement was behind Milosevich's politics and held that NATO's request for the Hungarian involvement to be nothing more than a mere intrigue. "We had been nearly forget that we had any problems with the Serbs, but now a demand of a great-power [...] drives us into this war that is not even a war yet." (Csurka, MIÉP: 29/17) Thus, refusing NATO's request appeared as a question of integrity on the side of Hungarian government. "no matter how small we are, there should be no way for ordering us about" (*ibid.*).

Regarding the two other parties in opposition, socialists were more critical about the Hungarian involvement, although ensuring their approval for the Decree from the beginning. They cast doubts on whether NATO together with Hungarian defence assured the security of Hungarians (on both sides of the border) to the necessary degree or not. They also emphasized "It is our conviction that the Kosovar conflict cannot be solved with military intervention..." (Kovács, MSZP: 21/17) They stressed the role of UN and the international negotiators too, besides that of the NATO. The liberal SZDSZ was emphasizing in turn the role of human rights, minority rights and the right for human dignity. "[It is our interest that] ...nobody should think of stretching his powers without any limitations, or think that he can do anything with the citizens or minority [of the country] neither in our region nor elsewhere in Europe" (Szent-Iványi, SZDSZ: 25/17). They thought NATO to be a political and valued community that acted for European security, peace and stability. The liberals did not share the worries of MSZP about Hungary's security. SZDSZ drew a distinction between the Serb nation and Milosevich.

In response to questions raised, prime minister Orbán pointed out that Hungary is not involved in the first stage of the planned intervention and the regions of Yugoslavia where Hungarians live were not targeted. He underlined the role of the government in arriving at these decisions within NATO. In answering to the repeated questions of the socialists, finally he stated that he believes Hungary can protect all of its citizens with the help offered by NATO, but added "in closed committee meetings all classified military information necessary for the informed choice were shared with the party fractions. I ask all of you not to shift the responsibility of your personal decision onto me, onto the cabinet or onto the coalition." (Orbán, prime minister: 85/17) Concerning the Hungarian minority in Yugoslavia, he raised the question "if the Hungarians living in Vajdaság will experience insults might we rely more on the necessary help of the international community if we don't participate in this intervention or not?" (*ibid.*).

The number of serious questions raised by the opposition and the dilemmatic answers by the government reflected the complexity and ambivalent nature of the Hungarian involvement in the conflict. As the foreign minister in his response repeated, apart from Hungary's future NATO membership, assurances, and the problems of the Hungarian minority in Yugoslavia, this decision is about values, Hungary cannot give the cold shoulder to the severe violations of human and minority rights, the genocide and ethnic cleansing, the

humanitarian catastrophe. Standing up for these values and following the norms of the North-Atlantic community were thus important factors in reaching an almost consensual decision.

European security and the protection of minority rights were seen as closely connected in the Hungarian representation of the conflict. On the other hand, NATO was portrayed as an alliance similar to binding international coalitions that Hungary was member in the 20th century (e.g. 1st and 2nd World War, Warsaw treaty) by MIÉP. The prime minister's response to these historical analogies made an important distinction, that "this alliance was chosen by ourselves for the first time in Hungarian history". He added that "the smaller a country the more it gets to situations in which there are raised questions to be answered [...] this does not mean however that we cannot choose the answer to those questions, now we are about to make this choice." (Orbán, prime minister: 31/17). Fortunately, there was no need for military interventions in autumn 1998.

Amending initial contributions, changes in spring 1999

When in spring 1999 the peace talks were blocked and NATO operations were decided upon, Hungary was already a member of the transatlantic organization. Thus, Article 5 of the Washington Agreement was applicable, providing a full security assurance to Hungary. For the implementation of this protection, an amendment to the original Decree was proposed on 24 March 1999 that allowed NATO aircrafts to land and depart from Hungarian airports for "military or technical" reasons using also ground service. After an exposé by the foreign minister Martonyi, a brief discussion was held where all sides repeated their arguments with slight modifications. Most party fractions regarded this amendment as an extension of the decision that has been made in October 1998. Liberals added that now the credibility of NATO is at stake and admitted that while similar conflicts might be unsolved throughout the world "we have to face this situation now" (Kuncze, SZDSZ: 140/58). Socialists drew attention to the possible migration and invited the government to take measures avoiding discontent in public feelings. FIDESZ advised that Hungary remain passive, and should not engage directly in military attacks. MIÉP adhered to its anti-NATO position revealing more and more their revisionist political goals as the conflict escalated.

Later, in the course of the NATO interventions important changes occurred: NATO targeted those northern regions of Yugoslavia that were inhabited by the Hungarian minority, and in the escalating conflict Hungarian men were forcibly recruited to fight in the Yugoslav army in Kosovo. This further motivated members of the Hungarian minority together with other refugees to flee to Hungary. After the intensification of air strikes the possibility of ground operations was also considered and Hungarians were concerned about the implications of these operations.

On 12 April, 1999, prime minister Orbán updated parliament regarding the state of the war and about the Hungarian involvement. He started by expressing his gratitude for the consensual support of the government policies within the Parliament, noting that it is not the case in every country of the region. He then tried "to shed light on this many faceted, peculiar international situation" (Orbán, prime minister: 4/61). NATO, he said, does not want military intervention from Hungary, a Hungarian medical team was to be sent to help the refugees in Albania, and thirdly, Hungary cannot be successfully attacked. He mentioned the case of a Russian humanitarian convoy that was stopped at the Hungarian border because it included armed vehicles and was carrying too much fuel. An agreement was reached between Hungarian and Russian authorities. The most interesting problem for Hungary was the future of the Hungarian minority. He said, this "makes the current international situation difficult and complex for Hungarians" (*ibid.*). Hungary is interested in ending this conflict as soon as possible. The speaker of FIDESZ repeated that they are convinced that NATO is defending

the democratic rights that connect the members of the organization and that the intervention is moral as it stops ethnic cleansing. SZDSZ, expressing their support for both the NATO intervention and cabinet actions, pointed out that the conflict could be lost politically even if there is no question about the military victory. They urged the government to play an active role in the settlement after the conflict. The socialists agreed that Hungary must fulfil its obligations to its allies, but “with undertakings that are associated with no more risk than necessary” (Kovács, MSZP: 10/61). They urged the government to be more active even in brokering an agreement in the conflict, relying on “our” rich experience about the Balkan. MIÉP attacked the Serb minority in Hungary together with the communist party (MSZMP) for demonstrating against NATO intervention. They also expressed that “We have only one duty, and we regard only this: what is the future for Hungarians?” “The whole of Europe will have to be rearranged after this conflict. [...] this is the end of the Versailles-system⁷” (Csurka, MIÉP: 12/61). The foreign secretary in his response assured that the government is working to develop a detailed strategy for South-East Europe, defends Hungarian interests within NATO, and supports all strategic solutions that are developed by the Hungarian minority in Yugoslavia. He also assured that humanitarian aid would be sent to them before the end of the conflict.

During May-June, close to the end of the military conflict, one of the parties proposed a significant change in the Hungarian involvement. Socialists wanted to restrict Hungarian assistance to NATO by excluding the possibility of air strikes initiated from Hungary. They held that this was only to prevent Hungary from becoming more entangled in the conflict, and did not withdraw their general support. In explaining their position (01-June-1999), they were the first among moderate parties to comment upon NATO intervention in the Hungarian Parliament: “NATO’s military action did not reach the desired effects neither in military nor in political terms.” (Tabajdi, MSZP: 288/74) They justified their positions by the fact that “Within NATO we can see a variety of positions and different national interests appear.” (*ibid.*). Fischer’s and Amato’s proposals for cease fire, the Greek attempts in mediation, and the Czech Republic’s low-keyed participation in the NATO intervention were mentioned as examples. They were also motivated by the bombing of Vajdaság (Voivodina) and by the “growing popular concerns about the continuing war” (*ibid.*). None of the parties supported the attempts of MSZP. FIDESZ and MIÉP interpreted it as intermixing Hungary’s national interests with domestic concerns in contesting the government. The proposal of MSZP was not accepted for discussion, and the issue divided even the socialist faction.

The end of the conflict was announced in the Parliament by foreign minister Martonyi on 14 June, 1999. He emphasized, “These are new times for the whole world. This was the first war fought not for market, for resources, for territories [...] but for defending principles and values.” (Martonyi, foreign minister: 2/77) In this new situation “we should not move borders, but change their function radically in accordance with goals derived from Central European identity, cooperation between neighbours, and European integration.” (*ibid.*). The socialists claimed that they are “just as happy for peace as the government is” (Kovács: 22/77) It is a first step in a long way of stabilizing the region. The speaker for FIDESZ was satisfied that Hungary had become a sound partner of NATO, while the Hungarian minority had remained to be in relative security. It was only MIÉP that rejected the foreign minister’s interpretation. They held that NATO lost the war in a degrading and silly way when they “let the Russians into the negotiations” (Csurka: 26/77). They saw financial motives behind the intervention, thus did not accept that it was a war fought for values and principles. The only thing they accepted was the need for Central European co-operation.

⁷ This is a reference to the peace after the First World War.

Conclusions

In this example of Hungarian discourse it could be seen that a consensual basis was built around a core of seeing Euro-Atlantic integration as reinforcing democratic change within the region. This was regarded as providing a chance for ameliorating the tensions between having a significant number of national minorities and a rising state-nationalism (reviving 19th century concepts of nation, see Kosáry, 2001) in the region. European integration in the long run (when all neighbouring countries will be members) was seen as a peaceful transformation easing the tensions between nations and helping the problem of minorities by changing the role of national borders as well as the 19th concept of nationhood associated with a given territory of a nation-state. But political parties weighted and structured the three foreign policy priorities (and the notions of nation, EU, Europe) differently, putting them into different contexts (of values or ideology).

A dominant representation emerging from Hungarian political discourse thus sets a new kind of dynamics into the parallel processes of regionalisation and European integration studied elsewhere. Hungarian political elites see themselves as helping national integration while also keeping nation-states as political entities. However, liberals seemed to question this representation by giving priority to (individual) freedom against any form of collectives. From another perspective, the extreme right also contested it by maintaining the supremacy of nations and old-fashioned nation-states.

By accepting the values of the Transatlantic community the Hungarian political elite appeared to have an agreement transcending ideological and party differences. It was only the right wing MIÉP that interpreted the participation of the NATO intervention in terms of historic analogies of outside domination and Hungarian defencelessness.

Intergroup and international complexities did not disappear even in a heated situation of armed conflict. Social psychologists tend to regard intergroup relations as mainly characterized by the dichotomy of in- and out-groups. The representation of the situation presented here was far from being black and white. Dilemmas on the side of the government and rapidly growing ambiguity in the position of the opposition were both visible. Political parties' general strategies appeared in discussing and making political decisions in response to challenges and crises. The Kosovar conflict created a dilemma for the Hungarian political elite. Values or political priorities of all political parties were not easily applied to this complex situation. The governing right-wing coalition, which had been distancing itself more from the EU and emphasized the Hungarian interests more, stood firm, while the pro-European socialists became hesitant in supporting NATO. In the changing position of the latter the growing worries of the general public could be reflected more freely than in the position of the right wing that had to cope with government responsibility.

Parliamentary discourse in this case was followed by significant actions. Therefore it was interesting to see how the original, almost unanimous, support dissipated over time and reached a phase of challenging the government's positions. Had the conflict being further intensified the opposition might have had formulate its counter proposals in a way that more explicitly broke the consensus over the support of the NATO intervention.

In sum, in this example of parliamentary discourse we could see how overall representations of international relations, their historic change, and the national perspective (popular aspirations) may determine the views of political actors and ultimately their behaviour. It was also important to see that even in such a heated situation the opposing parties positions could not be characterized in black and white terms and the shades changed dynamically over time.

References

- Canter, D.V. & Monteiro, C. (1993) The lattice of polemic social representations: A comparison of social representations of occupations in favelas, public housing, and middle class neighbourhoods of Brazil. In G. M. Breakwell & D.V. Canter (Eds.) *Empirical approaches to social representations* (pp. 223-247). Oxford: Clarendon Press.
- Csizmadia E. (2001). *Diskurzus és diktatúra. A magyar értelmiség vitái Nyugat-Európáról a késő Kádár-rendszerben. [Discourse and dictatorship. Debates of the Hungarian Elites at the end of Kádár's system]* Budapest: Századvég Kiadó.
- Doise, W. (1993) Debating social representations. In G. M. Breakwell & D.V. Canter (Eds.) *Empirical approaches to social representations* (pp. 157-170). Oxford: Clarendon Press
- Doise, W., Clémence, A. & Lorenzi-Cioldi, F. (1993) *The quantitative analysis of social representations*. London: Harvester Wheatsheaf.
- Doise, W. & Devos, T. (1999). Identité et interdépendance: pour une psychologie sociale de L'Union Européenne. *Psychologie et Société, 1*, 11-27.
- Granasztói, Gy. (2000). Erkölcs és politika európai csatlakozásunk küszöbén. *Magyar Szemle [Hungarian Review]*, 9(9-10), 5-24.
- Kiss P. (2001) Személy- és csoportpercepció egy nemzetközi konfliktusban: a megfigyelő perspektíváját meghatározó laikus elméletek szerepe. [Person and group perception in an international conflict: role of lay theories influencing perceiver's perspective] In Gy. Hunyady & L.A. Nguyen Luu (Eds.), *Sztereotípiakutatás. Hagyományok és irányok* (pp. 417-436). Budapest: Eötvös Kiadó.
- Kiss, P. (2002). Social representations of Europe and neighbouring nations in the Carpathian Basin. *Applied Psychology in Hungary*, 3-4, 65-79.
- Kiss, P. (2005). Hungarians about their Nation, Others, and Europe: Identification and representation processes in public. *Applied Psychology in Hungary*, 5-6, 85-102.
- Kiss P. & Hunyady Gy. (2005). Nation, State and National Identity in Modern Hungary. In W. Spohn & A. Ichijo (Eds.), *Entangled identities: Nations and Europe* (pp. 119-136). Aldershot: Ashgate Publishing Ltd.
- Kosáry, D. (2001). *A magyar és európai politika történetéből*. [From the History of the Hungarian and European politics] Budapest: Osiris.
- Kovács, M. (2001). Elmarasztalni és elhalasztani. Az EU diszkurzív stratégiái az 1998-as és 1999-es országjelentésekben [Condemning and postponing. EU's strategies in the 1998 and 1999 country reports]. *Replika*, 45-46, 89-110.
- Levy, J. S. (2003). Political psychology and foreign policy. In D. O. Sears, L. Huddy & R. Jervis (Eds.), *Oxford Handbook of Political Psychology* (pp. 253-284). New York: Oxford University Press.
- Moscovici, S (1984) The phenomenon of social representations. In S. Moscovici & R. Farr (Eds.) *Social representations* (pp. 3-69). Cambridge: Cambridge University Press.
- Reicher, S., Hopkins, N. & Condor, S. (1997). Stereotype construction as a strategy of influence. In R. Spears et al (Eds.), *The social psychology of stereotyping and grouplife* (pp. 94-118). Oxford: Blackwell.
- Szabó, M. (2000) Politika vs. Politikai: Carl Schmitt „das Politische“ fogalma. [Politics vs. political: Carl Schmitt's concept "das Politische"] In M. Szabó (Ed.), *Beszélő politika. A diszkurzív politikatudomány teoretikus környezete* (pp. 24-50). Budapest: Jószyöveg Műhely Kiadó.
- Tajfel, H. (1981) *Human groups and social categories: Studies in social psychology*. Cambridge: Cambridge University Press.

A study of sociolinguistic variables in the British National Corpus

Kaoru Takahashi

Toyota National College of Technology

Abstract

This paper deals with the study of sociolinguistic variables which are fairly systematically associated with age, sex and social class. These variables are dealt with as linguistic features assigned to each speaker in each text in the spoken part of the British National Corpus. By means of multivariate analysis, the variation of the occurrence of selected linguistic features among registers will be classified. A multivariate analysis of this sort holds out the promise of being able to systematize domains in the corpus while also revealing the characteristic linguistic features of the groups classified. That is, the aim of an analysis of these variables is to yield a more sophisticated classification, based on the similarities and dissimilarities between register categories and also to characterize register variation in terms of linguistic style.

Keywords: subcorpus, age, sex, linguistic style, multivariate analysis, axis

Introduction

Trudgill (2002, p. 161) claims that, in terms of sociolinguistics, styles are characterized as varieties of a language which can range on a continuum from very formal to very informal. His notion of style is closely linked to so called factors, which constitute a speech community. Yet the notion of style cannot be limited by formality or informality alone. It may be possible to explore other parameters of style regarding text, linguistic situations and communities. Such an exploration may redefine the notion of style. It is my intention in my study to expand the sociolinguistic notion of style by the inclusion of a number of other factors in the exploration of style and social contextual variables.

In this paper, I will focus on the notion of style by considering prestige and vernacular forms. Swann (1996, p. 310) claims these are largely associated with speech communities in that prestige varieties of English are associated with speakers and listeners with high socio-economic status and with settings and topics best characterized as formal. Vernacular varieties are set in opposition to these at the other end of the spectrum. Labov (1972) uses the term vernacular to refer to the least self-conscious style of speech used by people in relaxed conversation with friends, peers and family members, so that the vernacular style represents informal speech oriented towards a local community. Although it may be modified under the influence of more public-oriented interaction, e.g. educational settings, media language and the influence of other social groups, we can safely say that the vernacular style remains the basic style for some groups. In fact, the term vernacular is used in a number of ways. As

Holmes (1992, pp. 80-81) claims, in a multilingual speech community, the many different ethnic or tribal languages used by different groups are referred to as vernacular language and the term vernacular generally refers to the most colloquial variety in a person's linguistic repertoire. In other words, the vernacular is the variety used for communication in the home and with close friends. Holmes (1992, p. 81) also claims that the term vernacular is sometimes used to indicate that a language is used for everyday interaction, without implying that it is appropriate only in informal domains.

In the light of both gender and social class, Labov (1966) in his study of New York City found that within each social class group, and across each stylistic context, the female informants tended to use more prestige or high-status language features, while the male informants used more vernacular language features. According to Holmes's (1992) work, an American study compared the speech of women in service occupations, working in garages and hotels, with the speech of women working at home. Those in paid employment used more standard forms than those working at home. It is obvious that those working at home have a preference for vernacular forms. When we focus on the factor of sex again, Holmes's (1992, p. 166) claim is interesting:

"Sex differences in language are often just one aspect of more pervasive linguistic differences in the society reflecting social status or power differences. If a community is very hierarchical, for instance, and within each level of the hierarchy men are more powerful than women, then linguistic differences between the speech of women and men may be just one dimension of more extensive differences reflecting the social hierarchy as a whole."

However, these notions are nothing but a claim and how to substantiate them in terms of lexis or grammar depends on the design of any given study.

I have been exploring in more detail the social variables that have been considered in research on speech communities to date. In this paper, especially, I wish to demonstrate how social variables such as sex and age may characterise a speech community. In doing so I will show how studies of this sort lead to the notion of prestige and vernacular forms.

Method

Overview of the method

The BNC World Edition is a 100 million word collection of samples of written (82.82%) and spoken (17.78%) British English from the late 20th century. Spoken texts are organised in two parts, that is, a context-governed part containing orthographic transcriptions of recordings made at specific types of meeting and event, and a demographic part containing orthographic transcriptions of spontaneous natural conversations made by members of the public. The present study addresses the demographic part only. This contains 4,211,216 words.

Firstly, I will focus on age variable encoded in the corpus metadata. This paper also deals with the cross-sectional distribution between two variables of the spoken part: sex variable and age variable because we can expect that the analysis by two variables can reveal finer variations of subcorpora. In order to analyse the metadata I will employ a multivariate technique, hence the similarity and the dissimilarity in the axes by the variables are made explicit. By identifying salient part-of-speech tags as markers of classified subgroups, then, my study proceeds to the next step in which the relevant words and lexis are focused on.

On the BNC

In the demographic part of the BNC, the approach of the selection method adopted uses demographic parameters to sample the population of British English speakers in the United Kingdom. In this approach, 124 adults over the age of 15 were recruited from across the United Kingdom as respondents. The approach is controlled in terms of three selection criteria, age (sdeage), gender (sdesex) and social group (sdecla).

To begin with, in order to identify the distribution of word frequencies of each cell produced by attributes of both sdeage and sdesex, the word frequencies are counted and then displayed in Table 1.

Table 1
The distribution of word frequencies of subcopora by adeage and adesex

		sdesex0	sdesex1	sdesex2	sdesex %
		Unknown	Male	Female	
sdesex %		0.38	41.14	58.47	
sdeage1	0-14	100.00	7.73	4.69	6.31
sdeage2	15-24	0.00	9.98	19.84	15.71
sdeage3	25-34	0.00	15.47	23.60	20.16
sdeage4	35-44	0.00	19.37	20.51	19.96
sdeage5	45-59	0.00	22.85	22.85	22.76
sdeage6	60+	0.00	24.60	8.50	15.09
sum %		100.00	100.00	100.00	100.00

Note. The grammatical function of each tag is described respectively when the tag appears.

According to this table, Unknown data is restricted to the under 14 age group. As compared with the distributions of both sexes, it is observed that the older the respondent is, the larger the text size becomes for males. However, for females the youngest and the oldest group are relatively small in size. Nevertheless, the distribution of frequency among cells subcategorised by both sdeage and sdesex is not heavily biased.

On the multivariate technique

Multivariate analysis is a statistical procedure concerned with the analysis of multiple measurements of each individual or object in one or more samples. The technique is used in various areas such as psychology, sociology and biology. In linguistics, due to the availability of electronic texts, multivariate analysis promises to allow a computer to do more than just count the frequencies of words or find the functions of specific features in the texts.

For the moment, I will consider the history of this approach. Originally, this approach was used to characterize and compare varieties of use in English language corpora (see Nakamura, 1994; Tabata, 1995; Takahashi, 1996; Takahashi, 1999; Lee, 2000; and Takahashi, 2006). Recently, Conrad and Biber (2001) edited a collection of relevant studies, which includes several studies (see Atkinson, 2001; Finegan, 2001; Connor-Linton, 2001; Conrad, 2001; Biber & Finegan, 2001; Connor-Linton & Shohamy, 2001; Reppen, 2001; Biber, 2001a; and Biber, 2001b). Not all the studies are concerned with the multi-feature and multi-dimensional approach and it is apparent that, despite these studies, little is still known about the essence of the multi-feature and multi-dimensional approach.

My study employs the Extended Hayashi's Qualification Method Type III (hereafter EHT3). This type of method is widely known as correspondence analysis. Similar techniques were developed independently in several countries, where they were known as optimal scaling, optimal scoring or homogeneity analysis. The purpose is to find the similarities and dissimilarities in the frequency matrix between social variables and between tags as well.

Table 2 is a schematic table of frequency matrix among attribute, age, concerning age categories and some tags in the BNC.

Table 2
A schematic table of frequency to one sentence in each category

sdeage	NN1	PUN	AT0	PRP	AJ0	NN2	PNP	...
1	0.497	1.171	0.279	0.265	0.193	0.121	1.011	...
2	0.483	1.171	0.290	0.284	0.221	0.119	1.028
3	0.539	1.117	0.336	0.296	0.219	0.127	1.063
4	0.498	1.103	0.301	0.276	0.197	0.123	1.025	...
5	0.521	1.102	0.344	0.307	0.214	0.128	1.082	...
6	0.551	1.108	0.374	0.334	0.234	0.146	1.086	...

Note. NN1: Singular common noun PUN: Punctuation, i.e., !,.,;or? AT0: Article
PRP: Preposition AJ0: Adjective NN2: Plural common noun
PNP: Personal pronoun

The procedure changes the positions of rows and columns of the original data matrix so that the proportionally large frequency figures can converge around the diagonal. Consequently, the social variables placed close together and the tags placed close together are considered to be qualitatively similar. Those located distant from one another are qualitatively different. Statistically, this analysis is based on the idea that it gives the quantities to social variables and tags in the original data matrix in the way that the given quantities yield the highest correlation coefficient between the two. As a result of this calculation, some dimensions or axes appear which serve as the criteria to characterise age groups. The dimensions are later interpreted sociolinguistically. Because both social variables and tags have weights on every dimension simultaneously, the weights can help us to interpret each criterion to characterise the age groups by considering the variations and relationships among age groups and parts of speech tag. It is also worth examining lexis relevant to the salient features, so features which are addressed in reference to lexical items are clarified later and discussed semantically.

Results and Discussion

Analysis of sdeage

Let me start the discussion of sdeage. Tables 3 and 4 are schematic tables of loading scores of the positive side and the negative side of axis 1 by EHT3 respectively.¹

Table 3
Loading scores of ages and tags on the positive side along axis 1 in sdeage

¹ Proportion accounted for indicates the percentage of the information in the original data explained to the axis in question. In this case, axis 1 is 0.6614, axis 2 is 0.1513 and axis 3 is 0.1001.

Categories		Positive side		
		Tags		
		VHN	The past participle form of the verb HAVE: <i>had</i>	0.23
		VHD	The past tense form of the verb HAVE: <i>had, 'd</i>	0.21
		VBN	The past participle form of the verb BE: <i>been</i>	0.18
		VHG	The <i>-ing</i> form of the verb HAVE: <i>having</i>	0.11
		EX0	Existential <i>there</i> , i.e., <i>there</i> occurring in the <i>there is ...</i> or <i>there are</i>	0.11
		PRF	The preposition <i>of</i>	0.11
		CJT	The subordinating conjunction <i>that</i>	0.11
		VBG	The <i>-ing</i> form of the verb BE: <i>being</i>	0.10
		UNC	Unclassified items	0.10
		CRD	Cardinal number	0.09
		VBD	The past tense forms of the verb BE: <i>was</i> and <i>were</i>	0.08
S6 (60+)	0.08			
		VHI	The infinitive form of the verb HAVE: <i>have</i>	0.07
		AJC	Comparative adjective (e.g. <i>better, older</i>)	0.07
:				
S5(45-59)	0.03			
:				
S3(25-34)	0.01			

Table 4
Loading scores of ages and tags on the negative side along axis 1 in sdeage

Categories		Negative side		
		Tags		
		PNQ	<i>Wh</i> -pronoun (e.g. <i>who, whoever, whom</i>)	-0.17
		VVZ	The -s form of lexical verbs	-0.15
		AVQ	<i>Wh</i> -adverb (e.g. <i>when, where, how, why, wherever</i>)	-0.13
		DTQ	<i>Wh</i> -determiner-pronoun (e.g. <i>which, what, whose, whichever</i>)	-0.12
		DPS	Possessive determiner-pronoun (e.g. <i>your, their, his</i>)	-0.11
S1 (0-14)	-0.09			
		VBB	The present tense forms of the verb BE,	-0.09
		VDB	The finite base form of the verb Do: <i>do</i>	-0.09
:				
S2 (15-24)	-0.03			
:				
S4 (35-44)	0.00			

The loading score of each feature indicates the extent to which one can generalize from an axis to a particular linguistic feature, or the extent to which a given feature is a

representative of the axis underlying a dimension and each feature. These features with loading scores including age groups are listed in these tables. In order to arrive at a more convincing interpretation of the axis, this section presents bar graphs of scores for each axis which is shown in Figure 1, so that the visualisation will help in the discussion of the axis.

In each graph, bars are grouped as follows: age groups, four verb groups and other typical grammatical groups such as nouns, adjectives and so on. It is relatively easy to identify similar grammatical groups because in the BNC the tags consist of a base, which is very often followed by suffixes marking subclass and/or inflection (usually in this order). Above all, the distribution of verbs is the most noteworthy in that 6 classifications, i.e. present tense, past tense, *-ing* form, infinitive form, past participle form and *-s* form will help us to grasp the characteristics of the axes, especially, in terms of tense.

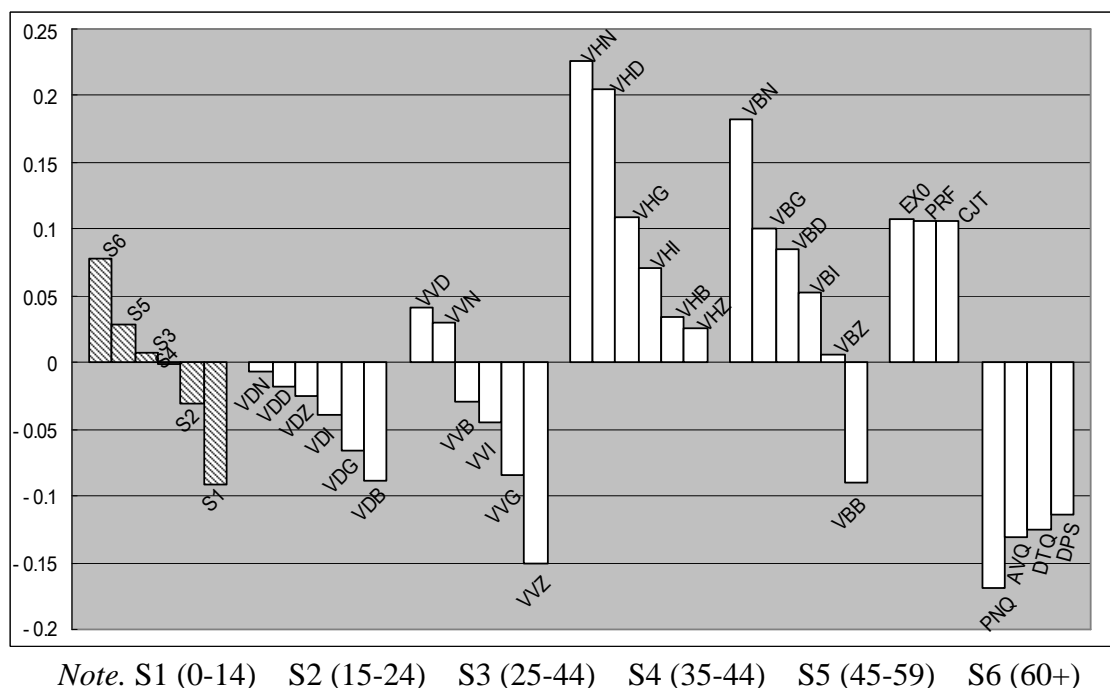


Figure 1. Bar Graph of Categories and Tags on axis 1 of sdeage

Taking a look at the bar graph, we should notice that the scores of EHT3 rank six age groups in the order of the age band along axis 1. Although there is an inversion between S3 (age 25-34) and S4 (age 35-44), this inversion is trivial enough to ignore. The higher the score of a feature is, the more relevant the feature becomes to the age of the speaker. Looking into the distribution of each verb group, the *do* verb group and lexical verb group tend to be more often used by younger people. In contrast, the *have* and *be* verb groups are more often used by older people. Moreover, the past participle forms of lexical verbs, *had* as a past participle form and *been* are typical markers of the older group, as well as the past tense form of the lexical verb groups, *had* as the past tense form, *was* and *were*. As for other features, the positive side relating to the older ages, EX0 (existential *there*, i.e., *there*), PRF (the preposition *of*) and CJT (the subordinating conjunction *that*) are salient markers; the negative side relating to the younger ages, PNQ (*wh*-pronoun), AVQ (*wh*-adverb), DTQ (*wh*-determiner-pronoun), and DPS (possessive determiner-pronoun). In order to interpret this axis, let me take a closer look at these tables and bar graph once again. What is notable on axis 1 of sdeage is the location of verb groups. Roughly speaking, the *do* verb groups and lexical verb groups are on

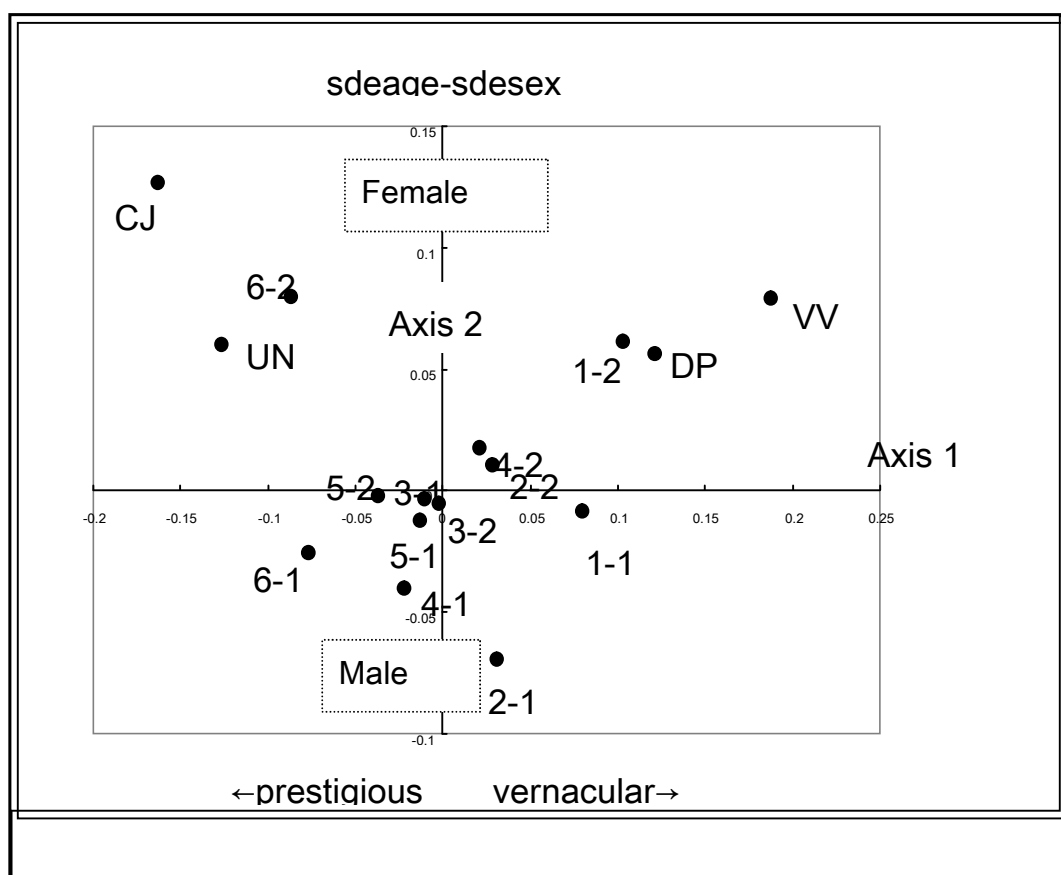
the side representing younger ages and the *have* verb group and *be* verb group are on the side representing older ages. The exceptions are VVD and VHN, on the side of the older ages, and VBB on the side of younger ages. On the side of the older ages, EX0, PRF and CJT are more commonly used than amongst the younger ages. AJC is also noteworthy because this tag does not appear as a salient feature. In contrast, for the younger ages, PNQ, AVQ, DTQ and DPS appear as salient features other than VVZ.

Now, let me discuss whether this axis is concerned with the notion of prestige and vernacular forms claimed by Labov (1966). That is, Labov uses the term vernacular to refer to the least self-conscious style of speech used by people in relaxed conversation with friends, peers and family members (see the Introduction section). The variable *sdeage* represents the age bands for demographic respondents, so if we suppose that the vernacular style forms the basic style for some groups, the young age group seems to be associated with this style. The relevant features are *wh*-pronoun, *wh*-adverb, *wh*-determiner-pronoun and possessive determiner-pronoun. Also, as for verbs, the *-s* form of lexical verbs, *be* and *do* are more often used by younger speakers, which may be markers of vernacular style. Furthermore, the consideration of the opposite side is noteworthy: the discussion of whether the group of older ages is associated with prestige style. Many features can be referred to, not only lexical features, i.e., unclassified items, cardinal numbers and comparative adjectives, but also grammatical features, i.e., existential *there*, the preposition *of*, and the subordinating conjunction *that*. Then in order to make my hypothesis more convincing, let me move on to the following section dealing with two social variables.

Analysis of subcorpora of sdesex-sdeage

This section addresses on the relationship between sex and age, whereby the axis appearing in this analysis gives us a hint to unresolved issue of prestige and vernacular forms hopefully. As the first step of the analysis, tag frequencies are counted in each cell of Table 1 and then EHT3 is carried out. In this case, the distribution of loading scores of subcorpora of *sdesex-sdeage* based on the scores of each cell is plotted in Figure 2.

As the result of the analysis, the subcorpora are also arrayed so as to rank age groups in the order of the age, roughly speaking. Given that the dimension or the axis which appeared in the previous analysis of age group is associated with the notion of prestige and vernacular forms, the dimension describes that the higher the loading score is, the more vernacular the subcorpora become and that, on the contrary, the lower the loading score is, the more prestigious the subcorpora become. If it is allowed to adapt this notion to an identical dimension, i.e., axis 1 of Figure 2, it is observed that there are differences between male and female, although a coherent pattern showing which sex is more prestigious among age groups is not identified. However, it is obvious that in age 35-44, the subcorpus of males is more prestigious than females. In other subcorpora, there is not so much difference. In any case, the discussion concerning prestige and vernacular forms is not far from persuasive so far, so the analysis shifts to the stage addressing lexis in the following section.



Note. combination of a-b

- a = 1. age 0-14
 2. age 15-24
 3. age 25-34
 4. age 35-44
 5. age 45-59
 6. age 60+

- b = 1. Male
 2. Female

Figure 2. Distribution of subcorpora by sdeage and sdesex, based on loading scores

Analysis in terms of lexis

In this section, focusing on the analyses in terms of lexis and semantics, we would be able to surmise that the appropriacy of the interpretation concerning prestigious style and vernacular style is verified. In order to realise the analyses, sophisticated programming is needed to produce a frequency matrix of words in targeted subcorpora. For example, when I focus on words tagged as ITJ (interjection), the programming needs to produce the frequency matrix of a certain number of the highest frequency words tagged as ITJ among subcorpora of adeage ranging from S1 to S6. The number of words should be designated arbitrarily, e.g. 50, 100, 150 or 200. Furthermore, the programming needs to deal with words among subcorpora generated by two variables such as adeage and adesex. Actually, this software was developed by the project group which was funded by a Grant-in-aid for Science Research from the Japan Society for the Promotion of Science and the Ministry of Education, Science, Sports and

Culture in 2001-2003. In this software, by inputting both the designated number of words to deal with, and also the variable(s) to produce subcorpora, the frequency matrix of the frequency of words and the occurrence normalised by the ratio to one sentence among subcorpora are produced at the same time.

Now, it is time to discuss the notions of prestigious style and vernacular style semantically, which are revealed as the most powerful dimension, i.e., dimension 1 in the analysis of subcorpora generated by variables concerning age, sdeage and sex, sdesex. In this case, the distribution of subcorpora of sdesex-sdeage based on the scores of each cell is plotted in Figure 2 along with salient features, based on the loading score of subcorpora and features. As mentioned previously, Dimension 1 describes that the higher the loading score is, the more vernacular the subcorpora become and that, on the contrary, the lower the loading score is, the more prestigious the subcorpora become. In the first quadrant, two features are beyond a subcorpus of sdeage 1, age 0-14 and sdesex 2, Female (labelled as 1-2), i.e., VVZ and DPS (possessive determiner-pronoun, e.g. *your, their, his*). They can be regarded as typical features to characterize vernacular style and Female. In the second quadrant, a subcorpus of sdeage 6, age 60+ and sdesex 2, Female (labelled as 6-2) are examined, focusing on CJT (the subordinating conjunction *that*), which is beyond this subcorpus, and UNC (unclassified items which are not appropriately considered as items of the English lexicon), which is very close to the subcorpus.

In the case of DPS, *my, your, her, his, their, our, its* and *me* are used in a subcorpus of sdeage 1, age 0-14 and sdesex 2, Female in descending order. Words tagged as DPS in a subcorpus of sdeage 1, age 0-14 and sdesex 2, Female are enumerated in descending order, along with w-units and occurrences in Table 5.

Table 5
Occurrences of words tagged as DPS in a subcorpus of sdeage 1, age 0-14 and sdesex 2

Words	<i>My</i>	<i>your</i>	<i>her</i>	<i>his</i>	<i>their</i>	<i>our</i>	<i>its</i>	<i>me</i>
w-units	647	514	233	192	71	62	52	9
occur.	0.0376	0.0299	0.0135	0.0112	0.0041	0.0036	0.0030	0.0005

Notes. Occurrences are normalised by the ratio to one sentence.

For the moment, let me discuss the problematic area of the tag annotation in the BNC. In this respect, *me* should be not included in DPS. Looking into the usage of *me* in all the 9 sentences, it is obvious that they are mistakenly tagged as DPS instead of PNP in a subcorpus generated by age 0-14 and Female. The following example can be referred to.

...sounds good to you don't it?, you can buy me cigarettes every day <laugh> <laugh>
(KP2 685 Title: 11 conversations recorded by 'Carla' (PS513))

In any case, attention should be given to the tag annotation. Moving further on to the next step of the analysis, the occurrences are compared with those in another subcorpus which is the closest to the origin of coordinates, i.e., the one generated by sdeage 3, age 25-34 and sdesex 2, female (labelled as 3-2). It is because the one located closest to the origin of coordinates can be regarded as neutral, meaning that it does not have any particular characteristic along the dimension. In other words, that a feature is located closest to the origin of the coordinates means the feature has a loading score nearly equal to zero, meaning that the feature is neutral along the dimension.

Moving on to the analysis of a subcorpus of sdeage 3, age 25-34 and sdesex 2, Female, words tagged as DPS in this cell are enumerated in descending order, along with the occurrences in Table 6.

Table 6
Occurrences of words tagged as DPS in a subcorpus of sdeage 3,
age 25-34 and sdesex 2, Female

Words	<i>your</i>	<i>my</i>	<i>his</i>	<i>her</i>	<i>our</i>	<i>their</i>	<i>me</i>	<i>its</i>
Ratio	0.0259	0.0191	0.0115	0.0075	0.0041	0.0040	0.0031	0.0009

Also, Table 7 shows the ratios when the occurrence of each word in a subcorpus of sdeage 1, age 0-14 and sdesex 2 is divided by that of sdeage 3, age 25-34 and sdesex 2, Female, along with their w-units and occurrences.

Table 7
Word tokens and frequencies concerning DPS in a subcorpus generated by age 0-14 and
female, along with their ratios

	sdeage1 and sdesex 2		sdeage 3 and sdesex 2		Ratio
	Word tokens	Frequencies	Word tokens	Frequencies	
<i>its</i>	52	0.002738	74	0.000849	3.224971
<i>my</i>	647	0.034071	1544	0.017706	1.924263
<i>her</i>	233	0.01227	607	0.006961	1.762678
<i>your</i>	514	0.027067	2102	0.024104	1.122926
<i>their</i>	71	0.003739	322	0.003692	1.01273
<i>his</i>	192	0.010111	931	0.010676	0.947078
<i>our</i>	62	0.003265	335	0.003842	0.849818
<i>me</i>	9	0.000474	248	0.002844	0.166667

Note. The ratio means the score when frequency in sdeage1-sdesex 2 is divided by that sdeage 3-sdesex 2. Words are enumerated in descending order of the ratios. As the score of the ratio is higher, the difference in the occurrences of features between the two subcorpora becomes large. Although a log-likelihood ratio test is quite useful to compare the occurrence, it is not suitable in my study. It is because a log-likelihood ratio test addresses the frequency data in text length; on the other hand, my study largely depends on the occurrence in sentence length, that is, the normalised occurrence to one sentence.

As a result of the above analysis, *its*, *my* and *her* are identified as typical features, which are thought to be largely associated with vernacular style in female. Let me once again consider the characteristic of vernacular form. As Labov claims, the term vernacular is used to refer to the least self-conscious style of speech used by people in relaxed conversation with friends, peers and family members. Therefore, among *its*, *my* and *her*, *my* is convincing. *Her* is also commonly used in the conversation of females, which is also acceptable. However, *its* is inexplicable. To solve this, sentences involving *its* in this subcorpus are examined. As a result, it is made clear that *it's* is mistakenly transcribed as *its* from speech data, as examples below show:

...unless *its* in the side of my motor bike ...

(KST 2882: 12 conversations recorded by 'Margaret2' (PS6RG), age 0-14, Female)

...no, I don't know where that tool is I was given for Christmas<pause>its gone...

(KST 2880: conversations recorded by 'Margaret2' (PS6RG), age 0-14, Female)

In my view, 34 examples out of 52 are mistakenly tagged, so *its* is not dealt with in this discussion. To sum up, typical words in vernacular style and Female are *my* and *her*.

Let me move on to another typical marker, VVZ, which far exceeds the subcorpus generated by age 0-14 and Female. The occurrences of words tagged as VVZ are enumerated in descending order in Table 8.

At a glance, we can confirm that they are verbs describing the action of people or expressing their will. According to Greenbaum and Quirk's (1990) claim concerning verb classes, these words refer to activity verbs categorised as dynamic verbs, or verbs of inert perception and cognition categorised as stative verbs. The former refers to *goes, says, gets, comes, looks, gives, takes, put* and so on; the latter refers to *wants, thinks, knows, likes* and so on. The activity verbs agree with event verbs and the stative verbs agree with state verbs advocated by Leech (1971). Leech claims that most of these words are among the most frequent verbs typically used as event or stative verbs, so let me regard these words as very straightforward cases. In contrast, relational verbs categorised as stative verbs such as *remains, requires, contains* do not appear. It follows from this fact that so long as we can observe the top 200 words tagged as VVZ, they are straightforward verbs used in daily life.

Table 8
Word tokens and frequencies concerning VVZ in a subcorpus generated
by age 0-14 and female

Word types	Word tokens	Frequencies
<i>goes</i>	551	0.029908
<i>says</i>	93	0.005048
<i>gets</i>	65	0.003528
<i>comes</i>	57	0.003094
<i>looks</i>	57	0.003094
<i>wants</i>	29	0.001574
<i>thinks</i>	29	0.001574
<i>knows</i>	28	0.00152
<i>gives</i>	18	0.000977
<i>takes</i>	16	0.000868
<i>means</i>	15	0.000814
<i>likes</i>	15	0.000814
<i>needs</i>	14	0.00076
<i>puts</i>	14	0.00076
<i>sits</i>	14	0.00076
<i>starts</i>	13	0.000706

Word types	Word tokens	Frequencies
<i>keeps</i>	11	0.000597
<i>works</i>	10	0.000543
<i>loves</i>	10	0.000543
<i>happens</i>	9	0.000489
<i>seems</i>	7	0.00038
<i>depends</i>	7	0.00038
<i>runs</i>	7	0.00038
<i>picks</i>	6	0.000326
<i>tells</i>	5	0.000271
<i>sees</i>	5	0.000271
<i>sounds</i>	5	0.000271
<i>stands</i>	5	0.000271
<i>lives</i>	5	0.000271
<i>feels</i>	4	0.000217
<i>helps</i>	4	0.000217
<i>pays</i>	4	0.000217

In order to highlight the characteristics of the subcorpus generated by sdeage 1, age 0-14 and sdesex 2, female, the occurrences of these verbs are compared with those in another subcorpus which is the closest to the origin of coordinates, i.e., the subcorpus of sdeage 3, age 25-34 and sdesex 2, Female. Table 9 enumerates words in descending order of the ratios. It shows that *goes* is the most distinctive word, followed by *sits, puts, thinks, gives, starts, loves* and *gets*.

Table 9
Word tokens and frequencies of VVZ of subcorpora of sdeage1-sdesex 2
and sdeage 3-sdesex 2, along with their ratios

	sdeage1-sdesex 2		sdeage 3-sdesex 2		
Words	Word tokens	Frequencies	Word tokens	Frequencies	Ratio
<i>writes</i>	3	0.000163	1	0.000012	13.58333
<i>leads</i>	2	0.000109	1	0.000012	9.083333
<i>reads</i>	2	0.000109	1	0.000012	9.083333
<i>goes</i>	551	0.029908	338	0.004033	7.415819
<i>dies</i>	3	0.000163	2	0.000024	6.791667
<i>cuts</i>	4	0.000217	3	0.000036	6.027778
<i>sits</i>	14	0.00076	11	0.000131	5.801527
<i>puts</i>	14	0.00076	14	0.000167	4.550898
<i>includes</i>	1	0.000054	1	0.000012	4.5
<i>explains</i>	1	0.000054	1	0.000012	4.5
<i>agrees</i>	1	0.000054	1	0.000012	4.5
<i>manages</i>	1	0.000054	1	0.000012	4.5
<i>stands</i>	5	0.000271	6	0.000072	3.763889
<i>thinks</i>	29	0.001574	37	0.000442	3.561086
<i>picks</i>	6	0.000326	8	0.000095	3.431579
<i>sends</i>	3	0.000163	4	0.000048	3.395833
<i>runs</i>	7	0.00038	10	0.000119	3.193277
<i>asks</i>	2	0.000109	3	0.000036	3.027778
<i>pays</i>	4	0.000217	6	0.000072	3.013889
<i>gives</i>	18	0.000977	30	0.000358	2.72905
<i>stops</i>	3	0.000163	5	0.00006	2.716667

Note. Words with a score of more than 10 in w-units sdeage 1 and sdesex 2 are focused on.

When I examine the usage of *goes* in a subcorpus of age 0-14 and Female, a very distinctive characteristic is revealed. That is, compared the occurrence of *I goes* in this subcorpus with that in a subcorpus of age 25-34 and female, which is closest to the origin of ordinates, the former one is 0.0017; the latter is 0.00002, meaning that *I goes* is far commonly used in a subcorpus of age 0-14 and Female. 29 examples out of 548 can be extracted from this corpus. One example is below.

So I goes, you must be a pervert too and you started laughing.

(KPG 931 Title: 39 conversations recorded by 'Josie' (PS555), age 0-14, Female)

The -s in *says* is known as the conversational historical present. However, given that *I goes* is mainly used in age 0-14, it is simply due to syntactic error or it is a phrase intentionally used as part of relaxed conversation in a friendly atmosphere. Otherwise, *I goes* can be regarded as Scottish or Irish dialect.² Whichever it is, it is safe to say that *I goes* is a typical characteristic

² Labov's work in Philadelphia focused on divergence between black and white speech (see Labov & Harris, 1986, p. 17), as witnessed in innovations like the use of -s as the marker of the past in narratives, rather than the traditional third person singular of the present tense. In the case of zero of the -s, conversely, African American Vernacular English is evolving a vernacular rule of zero (i.e., no suffix for all persons) for the present tense and -s for narrative past (in all persons). Moreover, there is a sharp distinction between the third person singular form of the present tense regular verb (e.g. standard *she walks* vs vernacular *she walk*) between the middle class groups and the lower-class groups (see Trudgill, 1983, p. 44 and Wolfram & Fasold, 1974).

of vernacular styles. In the case of other salient features, i.e., *sits*, *puts*, *thinks*, *gives*, *starts*, *loves*, *gets*, which are regarded as typical verbs used in daily life, Holmes' (1992, p. 81) claims can be referred to, that is, the term vernacular is sometimes used to indicate that a language is used for everyday interaction, so these words are said to be typical features in vernacular style. The example sentences are as follows.

So he gives him a pint of water, he goes and sits with the other ones.

(KPG 2523 Title: 39 conversations recorded by 'Josie' (PS555), age 0-14, Female)

Jenny just sits there and Leanne.

(KBF 13346 Title: 39 conversations recorded by 'Josie' (PS555), age 0-14, Female)

...she gets the kitten and puts it on my face and the kitten sits there going <"imitates cat licking">get off me!

(KPG 4088 Title: 39 conversations recorded by 'Josie' (PS555), age 0-14, Female)

And he puts, he puts his arm round him and he goes like, he's got his arm round him and squeezing him like that.

(KPG 4889 Title: 39 conversations recorded by 'Josie' (PS555), age 0-14, Female)

Next, let me move on to the prestigious side, which refers to the negative side of Figure 2. Subcorpora of age 60+, whether Male or Female (labelled as 6-1 and 6-2, respectively), are typical ones in this side. However, a subcorpus in the second quadrant, that is, the one generated by sdeage 6, age 60+ and sdesex 2, Female (labelled as 6-2) alone is focused on because UNC (unclassified items which are not appropriately considered as items of the English lexicon) and CJT (the subordinating conjunction *that*), which are in this quadrant, help to characterize this subcorpus. Table 10 enumerates words or descriptions of UNC in descending order of occurrence.

Table 10
Word tokens and frequencies concerning UNC in a subcorpus of sdeage 6-sdesex 2

lexicon	Word tokens	Frequencies	lexicon	W. t.	Frequencies
<i>erm</i>	1387	0.047652	<i>t</i>	17	0.000584
<i>er</i>	1378	0.047343	<i>sh</i>	17	0.000584
<i>Erm</i>	895	0.030749	<i>d</i>	16	0.00055
<i>Er</i>	258	0.008864	<i>tha</i>	15	0.000515
<i>&formula;</i>	206	0.007077	<i>*</i>	14	0.000481
<i>q.v.</i>	89	0.003058	<i>I</i>	14	0.000481
<i>ai</i>	88	0.003023	<i>Le</i>	13	0.000447
<i>s</i>	53	0.001821	<i>int</i>	10	0.000344
<i>sa</i>	34	0.001168	<i>/</i>	8	0.000275
<i>des</i>	33	0.001134	<i>e</i>	8	0.000275
<i>wh</i>	31	0.001065	<i>c</i>	7	0.00024
<i>&bgr;</i>	29	0.000996	<i>de</i>	6	0.000206
<i>di</i>	25	0.000859	<i>ha</i>	6	0.000206
<i>q.v</i>	24	0.000825	<i>f</i>	6	0.000206
<i>&frac12;</i>	19	0.000653	<i>non</i>	5	0.000172
<i>+</i>	17	0.000584	<i>&agr;</i>	5	0.000172

Note. Lower cases and upper cases are counted separately in counting frequencies of words tagged as UNC.

According to this table, pause fillers such as *er* and *erm* are commonly used. It is also noted that &formula (mathematical formula) is relatively frequent. In order to identify whether they are salient in a subcorpus of age 60+ and Female, these occurrences are compared with a neutral subcorpus generated by age 25-34 and Female (see Table 11).

Table 11
Word tokens and frequencies of UNC of subcorpora of sdeage6-sdesex2 and sdeage3-sdesex2, along with their ratios

	sdeage6 and sdesex2		sdeage3 and sdesex2		
Lexicon	Word tokens	Frequencies	Word tokens	Frequencies	Ratio
<i>c</i>	7	0.00024	2	0.000025	9.6
<i>tha</i>	15	0.000515	5	0.000062	8.306452
<i>q.v.</i>	89	0.003058	57	0.000703	4.349929
<i>sh</i>	17	0.000584	11	0.000136	4.294118
<i>&bgr;</i>	29	0.000996	19	0.000234	4.25641
<i>f</i>	6	0.000206	4	0.000049	4.204082
<i>wh</i>	31	0.001065	21	0.000259	4.111969
<i>s</i>	53	0.001821	39	0.000481	3.785863
<i>ai</i>	88	0.003023	72	0.000888	3.404279
<i>Le</i>	13	0.000447	11	0.000136	3.286765
<i>di</i>	25	0.000859	22	0.000271	3.169742
<i>o</i>	4	0.000137	4	0.000049	2.795918
<i>e</i>	8	0.000275	8	0.000099	2.777778
<i>des</i>	33	0.001134	36	0.000444	2.554054
<i>t</i>	17	0.000584	19	0.000234	2.495726
<i>sa</i>	34	0.001168	39	0.000481	2.428274
<i>q.v</i>	24	0.000825	28	0.000346	2.384393
<i>erm</i>	1387	0.047652	1623	0.020027	2.379388
<i>er</i>	1378	0.047343	1689	0.020841	2.271628
<i>d</i>	16	0.00055	21	0.000259	2.123552

Note. The ratio means the score when frequency in sdeage6-sdesex2 is divided by that in sdeage3-sdesex2.

In fact, the distinctive words are *c*, *tha*, *q.v.*, *sh*, *&bgr;*, *f*, *wh*, etc. The *c*, for example, seems used as a pause filler as follows:

He c-- , he c-- he couldn't, it don't matter what he does, he could never win a bloody war.(KCP 2795 Titled: 26 conversations recorded by 'Joy' (PS0GM), age 60+, Female)

The same is true with *tha*, *sh*, *f*, *wh*, etc. That is, the vague words tend to be used by older females in order to continue on to the following sentence smoothly. Let me for the moment discuss the difference between fillers and hedges. Some corpus studies make a distinction between fillers and hedges in tag annotation. In the BNC, tags are assigned to each word, so hedges such as *you know*, *sort of* and *you see* are not identified as such. Chafe and Danielewicz (as cited in Biber, 1988, p. 240) state that the use of hedges in conversational discourse indicates an awareness of the limited word choice that is possible under the production restrictions of speech. Therefore, hedges are less specific markers of uncertainty in informal situations. If it is acceptable to confirm that hedges are more associated with vernacular forms than prestigious forms and similar to pause fillers in the use of conversation, there seems no doubt that pause fillers such as *c*, *tha*, *sh*, *f*, and *wh* are typical markers of

prestigious forms, although *Erm/erm* and *Er/er* has no distinction in use between older females and a neutral subcorpus.

Lastly, the other salient feature, CJT, can be focused on, as this is the typical feature that strongly reflects the characteristic of the second quadrant of Figure 2. Holmes (1992, p. 170) claims that in many speech communities, when women use a linguistic form more frequently than men, it is generally the standard form - the overtly prestigious form - that women favour. Therefore, CJT is a typical marker to indicate the prestigious form of sentences and also is largely associated with structural complexity and formal situations. To sum up, fillers and the subordinating conjunction *that* without deduction represent the characteristic of the prestigious forms and female. In this way, the relationship between the prestigious form and female is gradually made explicit.

Final considerations and directions for further research

In order to arrive at a more convincing conclusion concerning linguistic style, another strategy may be needed because the discussion of vernacular and prestigious style so far is not enough to come to the conclusion. Then next step is to focus on the implication with social class. In this respect, Holmes' (1992, p. 161) question is suggestive:

"The linguistic features which differ in the speech of women and men in Western communities are usually features which also distinguish the speech of people from different social classes. So how does sex interact with social class?"

Obviously, it is worth discussing the issue of social class. However, let me leave the discussion of social class untouched in this paper because a thorough discussion of social class need much more spaces to be described here. In any case, in the next step, I would like to reconsider how social attributes, i.e., sex and age and social class, operate in a society or situation in terms of linguistic style.

Lastly, a novelty in my study lies in the fact that a unique statistical technique is employed to interpret dimensions produced by the multivariate analysis. I could explore a way to interpret dimensions more convincingly. That is, the subcorpora are compared with one located closest to the origin of the coordinates in terms of their loading scores, on the grounds that the subcorpus located closest to the origin of coordinates can be regarded as neutral. By doing this, the linguistic features associated with tag annotation or lexis are highlighted as salient ones along the dimension.

My ultimate aim is to provide an account of style strikingly in tune with a continuum-based approach to dimensions focused on style and lexis. In this respect, thorough studies will be carried out in the near future.

References

- Atkinson, D. (2001). Scientific discourse across history: a combined multi-dimensional/rhetorical analysis of the Philosophical Transactions of the Royal Society of London. In S. Conrad & D. Biber (Eds.), *Variation in English* (pp. 45-65). London: Longman.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

- Biber, D. (2001a). Dimensions of variation among eighteenth-century speech-based and written registers. In S. Conrad & D. Biber (Eds.), *Variation in English* (pp. 200-214). London: Longman.
- Biber, D. (2001b). On the complexity of discourse complexity: A multi-dimensional analysis. In S. Conrad & D. Biber (Eds.), *Variation in English* (pp. 215-240). London: Longman.
- Biber, D. & Finegan, E. (2001). Intra-textual variation within medical research articles. In S. Conrad & D. Biber (Eds.), *Variation in English* (pp. 108-123). London: Longman.
- Connor-Linton, J. (2001). Author's style and worldview: a comparison of texts about American nuclear arms policy. In S. Conrad & D. Biber (Eds.), *Variation in English* (pp. 84-93). London: Longman.
- Connor-Linton, J. & Shohamy, E. (2001). Register variation, oral proficiency sampling, and the promise of multi-dimensional analysis. In S. Conrad & D. Biber (Eds.), *Variation in English* (pp. 124-137). London: Longman.
- Conrad, S. (2001). Variation among disciplinary texts: A comparison of textbooks and journal articles in biology and history. In S. Conrad & D. Biber (Eds.), *Variation in English* (pp. 94-107). London: Longman.
- Conrad, S. & Biber, D. (Eds.) (2001). *Variation in English*. London: Longman.
- Greenbaum, S. & Quirk, R. (1990). *A student grammar of the English language*. Harlow: Longman.
- Holmes, J. (1992). *An introduction to sociolinguistics*. London and New York: Longman.
- Labov, W. (1966). *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. and W. Harris. (1986). De facto segregation of black and white vernaculars. In D. Sankoff (Ed.), *Diversity and diachrony* (pp. 1-24). Amsterdam: Benjamins.
- Lee, D. (2000). *Modelling variation in spoken and written language: The multi-dimensional approach revised*. Ph.D. thesis, Lancaster University.
- Leech, G. (1971). *Meaning and the English verb*. Edinburgh: Pearson Education.
- Nakamura, J. (1994). Extended Hayashi's Quantification Method Type III and its application in corpus linguistics. *Journal of Language and Literature, University of Tokushima, 1*, 141-191.
- Reppen, R. (2001). Register variation in student and adult speech and writing. In S. Conrad & D. Biber (Eds.), *Variation in English* (pp. 187-199). London: Longman.
- Swann, J. (1996). Style shifting, codeswitching. In D. Graddol, D. Leith & J. Swann (Eds.), *English: History, diversity and change*. (pp. 301-337). New York: Routledge.
- Tabata, T. (1995). Narrative style and the frequencies of very common words: A corpus-based approach to Dickens' first person and third person narratives. *Journal of English Corpus Studies, 3*, 91-109.
- Takahashi, K. (1996). *Typology of registers in a written English corpus: Multi-feature and multi-dimensional analyses*. M.Phil. thesis, UMIST.
- Takahashi, K. (1999). A linguistic interpretation of the third powerful dimension in the LOB corpus. *Journal of English Corpus Studies, 6*, 33-65.
- Takahashi, K. (2006). A study of register variation in the British National Corpus. *Literary and Linguistic Computing, 16*(1), 111-126.
- Trudgill, P. (1983). *Sociolinguistics: An introduction to language and society*. Harmondsworth: Penguin.
- Trudgill, P. (2002). *Sociolinguistic variation and change*. Edinburgh: Edinburgh University Press.
- Wolfram, W. & Fasold, R. W. (1974). *The study of social dialects in American English*. Englewood Cliffs NJ: Prentice-Hall.

Fetishism and anxiety: A test of some psychodynamic hypotheses

Andrew Wilson¹

Department of Linguistics and English Language
Lancaster University

Abstract

A computer-assisted content analysis of a corpus of 676 German-language foot, shoe, and boot fetish fantasies using the Dresdner Angstwörterbuch demonstrated an elevated rate of anxiety themes when compared with a broad-based general corpus containing 500 samples of published written German. This lends some support to the psychodynamic theory that fetishism has its roots in anxiety. More specifically, rates of mutilation (i.e. castration) anxiety themes were significantly higher in the fetish texts than in the control texts, but no significant difference was detected between the two sets of texts in relation to separation anxiety. This finding seems to support the traditional Freudian theory of fetishism, as opposed to the theories of the object-relations school; however, a phrasal analysis of the corpus also supports a modification of the Freudian theory, in that the woman may be the fetishist's perceived *agent* of castration rather than its original victim.

Keywords: fetishism; psychoanalysis; castration anxiety; mutilation anxiety; separation anxiety; computer-assisted content analysis; Dresdner Angstwörterbuch; corpus linguistics.

Introduction

The psychodynamic traditions pioneered by analysts such as Freud, Jung, and Szondi offer a rich source of theories for understanding the deepest roots of both individual behaviour and cultural evolution. They have demonstrated numerous points of compatibility with older-established psychological frameworks (e.g. Plé, 1952; Nolan, 1958; O'Doherty, 1962), with contemporary neuroscience (e.g. Van de Vijver, 2000; Turnbull & Solms, 2007), and with evolutionary theory (e.g. Dediu, 1999; Hargitai, 2007). Many of them are also amenable to formal modelling, testing, and embedding within a self-organizing systems perspective (e.g. Langs, 1988; Goldstein, 1990; Mandell & Selz, 1995; Scandella, 2004; Wedemann, de Carvalho & Donangelo, 2008). However, the question arises: To what extent can we believe in some of their more basic constructs? Nowhere, perhaps, is this more often asked - even by professional clinical psychologists - than in the domain of sexual development, with its talk of concepts such as the Oedipus complex, the phallic woman, and castration anxiety. For instance, in a review of the impact of Freudian ideas on mainstream psychology, Westen (1998, p. 355) reflected on the history of his own attitudes to such theories:

¹ eiaaw@exchange.lancs.ac.uk

"Personally, I did not find any of Freud's psychosexual notions compelling until I began doing clinical work and found myself confronting phenomena that simply could not be easily explained by alternative theories. ... These were not simply instances of schema-driven assimilation, because I was actually quite resistant to several of these notions and initially found these observations more discrepant than confirming."

Part of the problem lies, perhaps, in the fact that much of psychodynamic theory has arisen from generalizing about reports of individual clinical case studies, some of them now more than a century old. In response to this situation, in recent years, there have been several calls for an increase in the empirical testing of psychodynamic constructs, going beyond the traditional qualitative analysis of case studies and also incorporating the use of non-clinical data (Barglow, Jaffe & Vaughn, 1989; Bornstein, 2005; Luyten, Blatt & Corveleyn, 2006; Bornstein, 2007).² These calls have tended to envisage experimental designs as the main alternative to case-study work, but the analysis of large samples (or "corpora") of naturally occurring texts is another scientific technique that can be used to complement active experimental work with informants (cf. Gilquin & Gries, 2009). Large specialized text corpora, gathered from the internet, are, indeed, increasingly being used as a source of psychological information (e.g. Fekete, 2002; Cohn, Mehl & Pennebaker, 2004). The psychological validity of such data is supported by the fact that predictions about aspects of human language processing, which have been made on the basis of broad-based corpora, appear to concur with the findings of neuroimaging and EEG studies (Mitchell, Shinkareva, Carlson et al., 2008; Murphy, Baroni & Poesio, 2009). Neuroimaging, in turn, is increasingly being used as another way of testing psychoanalytic theories (e.g. Beutel, Stern & Silbersweig, 2003). However, the advantage of corpus data over neuroimaging studies lies in the ease with which they can be collected and analysed and in the large sample sizes that can be obtained and handled. Although some authors have suggested that data gathered from or using the World Wide Web might be atypical as a consequence of social differences in internet usage (Skitka & Sargis, 2005), this assumption has not been supported by empirical studies (Gosling, Vazire, Srivastava & John, 2004).

Taking on board the criticisms and methodological suggestions reviewed above, the aim of this paper is to address empirically the relationship between sexual fetishism and anxiety, with the goal of resolving some doubts and disagreements between different schools of depth psychology.

Most psychodynamic theories of fetishism locate its roots in anxiety, but they disagree as to which types of anxiety are involved (Whitelaw, 1959; Wise, 1985). Probably still the most widely held view is that of Freud (1927), who argued that fetishism is a consequence of castration anxiety. In Freud's opinion, the fetish is a substitute for the female penis, which the child originally believes to be present but later discovers to be absent: to ward off the anxiety of losing his own penis, which is precipitated by this discovery in relation to the mother, the child creates the fetish as penis substitute so that he can maintain the illusion that she has not, in fact, been castrated at all. In contrast to this traditional Freudian view, another large group of theorists (within the object-relations school of psychoanalysis) see the fetish as having its origins in separation anxiety. The argument here is that the child fetishist has a strong desire to continue clinging to its mother and this desire then becomes transferred to those objects or body parts that once allowed for easy clinging (Clerk, 1983). Some have suggested that the fetish object is more specifically a substitute for the mother's breast, the original object of oral

2 An early example of this kind of approach is provided by Miller (1969) in relation to the Oedipus complex.

attachment (Wulff, 1946).

This study will use a large web-sourced corpus of German-language foot, shoe, and boot fetish fantasy stories to attempt to answer two research questions about these proposed relationships between fetishism and anxiety: first, do the fetish fantasy texts actually contain more anxiety themes than German-language writing in general, and, second, is there any evidence to support the view that fetishism has its origins in castration anxiety as opposed to separation anxiety?

Data and Method

The experimental data comprised 676 examples of German-language foot, shoe, and boot fetish fantasy stories, taken from a web site, now no longer in existence, which archived such stories on a continuing basis (<http://fussvideo.de.vu>). This was, and still remains, probably the largest single archive of amateur foot and shoe fetish fiction in existence, in any language. The data were collected in the winter of 2003 and comprised the entire story archive of that web site on the date of collection. The sample consists, in total, of 1,607,344 running words of text. The median individual text length is 1,440 running words with an interquartile range of 875 to 2,858 running words.

As a control sample of non-fetishistic writing, the study used the LIMAS Corpus, which is a large collection of published written German (Glas, 1975). Broad-based text corpora like LIMAS are widely used by researchers in both linguistics and other cognate disciplines (McEnery & Wilson, 2001; McEnery, Xiao & Tono, 2005; Lüdeling & Kytö, 2008). They are designed to be as representative as possible of the totality of language use, so that they can be used to make general statements about a language or serve as control samples in studies of particular types of language behaviour (Biber, 1993). The LIMAS Corpus is made up of 500 individual text samples, which represent some 33 different genres and subject domains such as fiction, press texts, law, science and technology, etc. The text categories that constitute the corpus are based on the main subject/genre headings of the *Deutsche Bibliographie* and the relative size of each category within the corpus is proportional to the number of publications listed under the relevant heading in the 1970-1971 edition of this work (Wolters & Kirsten, 1999). The LIMAS corpus contains 1,069,206 running words of text. The median text sample length is 2,121 running words with an interquartile range of 2,057 to 2,188 running words.

To identify the prevalence of different anxiety themes in the data, the texts were content analysed using the *Dresdner Angstwörterbuch*, or DAW for short (Berth, 2001; Berth, 2004). The DAW is a computer-readable lexicon which contains some 3,875 stem, word, and phrase entries, once it has been re-formatted for the relevant version of the computer software and duplicate entries have been removed. The entries in the DAW are classified into six subtypes of anxiety: death, mutilation, separation, guilt, shame, and diffuse. These are the categories defined by Gottschalk & Gleser (1969) for use in the manual content analysis of textual data, and they have undergone numerous clinical tests of construct validity involving biochemical as well as psychological and diagnostic variables (e.g. Gottschalk & Gleser, 1969; Gottschalk, 1979; Gottschalk, Lolas & Viney, 1986; Gottschalk, 1995). With regard to these categories, it should be noted especially that "mutilation anxiety is synonymous with 'castration' anxiety, and the descriptive items in the scale pertaining to this subtype of anxiety are derived from clinical psychoanalytic psychology" (Gottschalk, 1982, p. 275). In the present study, only the categories of mutilation and separation anxiety were used separately; however, a total anxiety score, based on the sum of all six categories in the DAW, was also calculated, to test whether the fetish texts contained more anxiety themes overall than the control texts.

In clinical practice, the content categories are normally applied to five-minute spontaneous speech samples in response to a prompt about an "interesting or dramatic life experience" (Gottschalk, 1982, pp. 286-288). However, their applicability to other types of verbal data, such as dream narratives and archival materials, has also been established (Gottschalk, Stone, Gleser & Iacono, 1966; Gottschalk, DeFrancisco & Bechtel, 2002).

The DAW has been cross-validated against the more widely used manual scoring instructions and DAW scores for all anxiety types have shown significant correlations ($p < 0.01$) with the manual scoring (Berth & Suslow, 2001). In Berth and Suslow's (2001) re-analysis of 152 manually scored texts, all of the correlations fell within the 0.5 to 1.0 range for a large effect size, and, in their analysis of a fresh set of 60 texts, four correlations were at this level, with a further two in the 0.3 to 0.49 range for a medium effect size (cf. Cohen, 1992).

The DAW was applied to the texts using the content analysis software package *CoAn für Windows* (Romppel, 2001). CoAn divides the input files into segments pre-marked by the analyst (in this case, the individual stories and corpus samples) and the words and phrases in these segments are then matched against the entries in the DAW. For each text segment, CoAn provides a frequency count of each DAW category, which shows how many word or phrase occurrences fell into that category. It also provides a word count showing the number of running words within each segment. From these figures, one calculates the following score for each anxiety type in each text segment (Berth & Romppel, 1999, p. 189):

$$\sqrt{\frac{100}{n}}(f + 0.5)$$

where f = the raw frequency of the anxiety type in the segment and n = the total number of running words in the segment. Total anxiety is calculated in the same way, except that here f = the sum of the raw frequencies of all categorized anxiety words in the segment.

Statistical comparisons between the two samples were made using the asymptotic Wilcoxon-Mann-Whitney rank-sum test, as implemented in the optional package *coin* for R for Windows 2.6.1 (Ihaka & Gentleman, 1996; Hothorn, Hornik, van de Wiel & Zeileis, 2008). An effect size r was computed using the following two equations (DeCoster, 2004, pp. 16 and 27), where ne = the size of the experimental group (the fetish texts), nc = the size of the control group (the LIMAS corpus texts), and z = the relevant z-score from the Wilcoxon-Mann-Whitney test:

$$g = z \sqrt{\frac{ne + nc}{ne(nc)}}$$

$$r = \sqrt{\frac{g^2(ne)nc}{(g^2(ne)nc) + (ne + nc)(ne + nc - 2)}}$$

A value of r between 0.1 and 0.29 indicates a small effect size, a value between 0.3 and 0.49 indicates a medium effect size, and a value between 0.5 and 1.0 indicates a large effect size (Cohen, 1992).

Results

Table 1 shows the medians and interquartile ranges for total anxiety and the two anxiety subtypes in the two samples. Table 2 shows the Wilcoxon-Mann-Whitney test results for comparisons between the two samples.

Table 1
Median anxiety scores for the fetish and LIMAS corpus texts
(interquartile ranges in brackets).

	Fetish texts	LIMAS corpus
Mutilation anxiety	0.488 (0.336-0.633)	0.340 (0.160-0.456)
Separation anxiety	0.309 (0.236-0.408)	0.337 (0.259-0.409)
Total anxiety	0.902 (0.739-1.065)	0.735 (0.552-0.934)

Table 2
Results of Wilcoxon-Mann-Whitney rank-sum tests for differences between the fetish and LIMAS corpus texts.

	z-score	p	Effect size (r)
Mutilation anxiety	13.07	< 0.0001	0.356
Separation anxiety	0.129	0.897	0.004
Total anxiety	9.631	< 0.0001	0.271

It will be seen from Table 2 that only one of the two anxiety subtypes showed a significant difference between the two samples. This was mutilation anxiety, which occurred at a significantly higher rate in the fetish texts than in the LIMAS corpus texts ($p < 0.0001$). The difference between the two samples demonstrated a medium effect size ($r = 0.356$). In contrast, there was no significant difference with respect to separation anxiety. Total anxiety was also significantly higher in the fetish texts than in the LIMAS corpus texts ($p < 0.0001$), with the difference here showing an effect size close to the borderline between a small and medium effect ($r = 0.271$; borderline = 0.299/0.300).

Discussion

The results of this analysis have shown that the foot, shoe, and boot fetish fantasy stories contained significantly more themes of mutilation (i.e. castration) anxiety than the set of control texts taken from the German written language in general. The overall rate of anxiety themes in the foot, shoe, and boot fetish stories was also significantly higher than in the control texts. However, there was no significant difference between the fetish stories and the control texts in the occurrence of separation anxiety themes.

The fact that the rate of anxiety themes overall was elevated in the fetish texts would seem to support the more general psychodynamic position that fetishism arises out of anxiety.

Although it may seem counterintuitive that the pleasant sense of sexual arousal to which the fetish gives rise should originate in the unpleasantness of anxiety, there is, in fact, ample empirical evidence that sexual arousal increases under conditions of anxiety, perhaps as a consequence of disinhibition (Dutton & Aron, 1974; Barlow, Sakheim & Beck, 1983; Palace & Gorzalka, 1990). From a semiotic perspective, the emergence of a fetish from the combination of anxiety and sexual arousal can be seen as yet another example of a third course behaviour, an important feature in human cultural evolution (Koch, 1993; Koch, 2009).

The findings on the subtypes of anxiety provide support for some more specific psychodynamic theories of fetishism, whilst raising doubts about others. On the one hand, the higher rate of mutilation anxiety in the fetish texts appears to lend support to the traditional Freudian theory that fetishism is a consequence of castration anxiety (Freud, 1927). On the other hand, the absence of any significant difference between the fetish and control texts in relation to separation anxiety would seem to pose problems for the competing set of theories in which the fetish object is deemed to be a breast substitute or other substitute for maternal clinging (Wulf, 1946; Clerk, 1983). This is not to deny that the fetish texts have a strong oral component: for instance, *in den Mund* (= "in my/his mouth") is the third most frequent three-word phrase in the fetish corpus, with 382 occurrences, whilst *mit meiner Zunge* (= "with my tongue") also occurs in the top twenty, with 235 occurrences (Wilson, 2009). However, in the absence of any significant difference in the degree of separation anxiety, it seems likely that these phrases reflect a form of oral stage regression in which the oral-sadistic phallic mother is seen as the source of potential castration (Stolorow, 1975; Creed, 1993; Borovečki-Jakovljević, Borovečki & Jakovljević, 2004), rather than as one of its victims, which has been the traditional position of Freudian theory: as Carter (2000, p. 83) puts it, "the fetish is a substitute for the maternal phallus, the fantasy-organ of the powerful and potentially castrating (rather than castrated) woman". Indeed, the frequency list of three-word phrases in the corpus of fetish stories also establishes that most references to the male genital region are associated with violent acts perpetrated by female characters, sometimes by the mother herself: for example, *Na, wie wäre das, von der eigenen Mutter mit hohen Stiefeln in der Eier getreten zu werden?* (= Well, what would it be like, to be kicked in the nuts by your own mother with tall boots?).

Conclusion

This empirical analysis of anxiety themes in a large corpus of German-language foot, shoe, and boot fetish fantasies has provided some support for the traditional Freudian theory of fetishism as arising out of castration anxiety, albeit in a modified form where the woman is seen as the source of potential castration rather than its original victim. In contrast, it has also suggested that the competing object-relations theory, based on separation anxiety, is not so transparently supported by the data of fetish fantasy stories, at least in relation to foot, shoe, and boot fetishism. However, it is worth remembering that the early sexologist Richard von Krafft-Ebing (1901) treated foot, shoe, and boot fetishism rather differently from other types of fetishism, noting that it was particularly closely linked with masochism. Later psychoanalysis has diluted this link, tending to treat fetishism and masochism as two separate entities with different aetiologies. Future research might care to explore whether the findings of this study, which dealt solely with foot, shoe, and boot fetishism, can be replicated in the case of other fetishes, and, if not, whether it is worth reconsidering Krafft-Ebing's separation of the fetishes in more detail.

References

- Barglow, P., Jaffe, C.M. & Vaughn, B. (1989). Psychoanalytic reconstructions and empirical data: Reciprocal contributions. *Journal of the American Psychoanalytic Association*, 37, 401-435.
- Barlow, D.H., Sakheim, D.K. & Beck, J.G. (1983). Anxiety increases sexual arousal. *Journal of Abnormal Psychology*, 92, 49-54.
- Berth, H. (2001). Die Messung von Angsteffekten mittels computergestützter Inhaltsanalyse. *Psychotherapie, Psychosomatik, Medizinische Psychologie*, 51, 10-16.
- Berth, H. (2004). *Das Dresdner Angstwörterbuch DAW: Entwicklung, Validierung und Erprobung einer Computerversion der Gottschalk-Gleser-Angstskalen*. Frankfurt am Main: VAS.
- Berth, H. & Romppel, M. (1999). Darstellung und Erleben der Wende in Massenmedien: Inhaltsanalytische Untersuchungen am Wendekorpus - zehn Jahre danach. *Medienpsychologie*, 11, 185-199.
- Berth, H. & Suslow, T. (2001). Zur Validität eines automatisierten inhaltsanalytischen Instrumentes der Angstmessung: Das Dresdner Angstwörterbuch. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie*, 49, 88-105.
- Beutel, M.E., Stern, E. & Silbersweig, D.A. (2003). The emerging dialogue between psychoanalysis and neuroscience: Neuroimaging perspectives. *Journal of the American Psychoanalytic Association*, 51, 773-801.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8, 243-257.
- Bornstein, R.F. (2005). Reconnecting psychoanalysis to mainstream psychology: Challenges and opportunities. *Psychoanalytic Psychology*, 22, 323-340.
- Bornstein, R.F. (2007). Nomothetic psychoanalysis. *Psychoanalytic Psychology*, 24, 590-602.
- Borovečki-Jakovljević, S., Borovečki, A. & Jakovljević, V. (2004). Phallic woman - myth or reality of our everyday life. *Socijalna Psihijatrija*, 32, 113-116.
- Carter, K. (2000). Consuming the ballerina: Feet, fetishism and the pointe shoe. *Australian Feminist Studies*, 15, 81-90.
- Clerk, G. (1983). Pour une lecture nouvelle du fétichisme. *Revue Française de Psychanalyse*, 47, 321-324.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Cohn, M.A., Mehl, M.R. & Pennebaker, J.W. (2004). Linguistic markers of psychological change surrounding September 11, 2001. *Psychological Science*, 15, 687-693.
- Creed, B. (1993). *The monstrous feminine: Film, feminism, psychoanalysis*. London: Routledge.
- DeCoster, J. (2004). *Meta-analysis notes*. Retrieved July 28, 2009, from <http://www.stat-help.com/notes.html>
- Dediu, D. (1999). *Evolutionary view of the Szondian Triebssystem*. Retrieved October 19, 2009, from <http://www.geocities.com/HotSprings/Villa/2928/hypotheses.html>
- Dutton, D.G. & Aron, A.P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, 30(4), 510-517.
- Fekete, S. (2002). The internet - a new source of data on suicide, depression and anxiety: A preliminary study. *Archives of Suicide Research*, 6, 351-361.
- Freud, S. (1927). Fetischismus. *Internationale Zeitschrift für Psychoanalyse*, 13, 373-378.
- Gilquin, G. & Gries, S.Th. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics & Linguistic Theory*, 5, 1-26.
- Glas, R. (1975). Das LIMAS-Korpus, ein Textkorpus für die deutsche Gegenwartssprache.

- Linguistische Berichte*, 40, 63-66.
- Goldstein, J. (1990). Freud's theories in the light of far-from-equilibrium research. In D. Andersen, G. Richardson & J. Sterman (Eds.), *System Dynamics '90: Proceedings of the 1990 International System Dynamics Conference* (pp. 440-454). Cambridge, MA: The System Dynamics Society (MIT).
- Gosling, S.D., Vazire, S., Srivastava, S. & John, O.P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59, 93-104.
- Gottschalk, L.A. (1982). Manual of uses and applications of the Gottschalk-Gleser verbal behavior scales. *Research Communications in Psychology, Psychiatry and Behavior*, 7, 273-327.
- Gottschalk, L.A. (1995). *Content analysis of verbal behavior: New findings and clinical applications*. Hillsdale, NJ: Erlbaum.
- Gottschalk, L.A., DeFrancisco, D. & Bechtel, R.J. (2002). Computerized content analysis of some adolescent writings of Napoleon Bonaparte: A test of the validity of the method. *Journal of Nervous and Mental Disease*, 190, 542-548.
- Gottschalk, L.A. & Gleser, G.C. (1969). *The measurement of psychological states through the content analysis of verbal behavior*. Los Angeles, CA: University of California, Berkeley.
- Gottschalk, L.A., Lolas, F. & Viney, L.L. (1986). *Content analysis of verbal behavior: Significance in clinical medicine and psychiatry*. Heidelberg: Springer Verlag.
- Gottschalk, L.A., Stone, W.N., Gleser, G.C. & Iacono, J.M. (1966). Anxiety levels in dreams: Relation to changes in plasma free fatty acids. *Science*, 153, 654-657.
- Hargitai, R. (2007). The Szondi test: Theory and practice. *Empirical Text and Culture Research*, 3, 85-105.
- Hothorn, T., Hornik, K., van de Wiel, M.A. & Zeileis, A. (2007). *Coin: Conditional inference procedures in a permutation test framework*. Retrieved September 21, 2009, from <http://cran.r-project.org/package=coin>.
- Ihaka, R. & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5, 299-314.
- Koch, W.A. (1993). *The roots of literature*. Bochum: Brockmeyer.
- Koch, W.A. (2009). *Tractatus logo-philosophicus: On the subquantum foundations of the world*. Norderstedt: BoD.
- Krafft-Ebing, R. von (1901). *Psychopathia sexualis: Mit besonderer Berücksichtigung der conträren Sexualempfindungen* (11th ed.). Stuttgart: Ferdinand Enke.
- Langs, R. (1988). Mathematics for psychoanalysis. *British Journal of Psychotherapy*, 5, 204-212.
- Lüdeling, A. & Kytö, M. (Ed.) (2008). *Corpus linguistics: An international handbook*. 2 vols. Berlin: Walter de Gruyter.
- Luyten, P., Blatt, S.J. & Corveleyn, J. (2006). Minding the gap between positivism and hermeneutics in psychoanalytic research. *Journal of the American Psychoanalytic Association*, 54, 571-610.
- Mandell, A.J. & Selz, K.A. (1995). Nonlinear dynamical patterns as personality theory for neurobiology and psychiatry. *Psychiatry*, 58, 371-390.
- McEnery, T. & Wilson, A. (2001). *Corpus linguistics* (2nd ed.). Edinburgh: Edinburgh University Press.
- McEnery, T., Xiao, R. & Tono, Y. (2005). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- Miller, A.R. (1969). Analysis of the Oedipal complex. *Psychological Reports*, 24, 781-782.
- Mitchell, T., Shinkareva, S.V., Carlson, A., Chang, K.M., Malave, V.L., Mason, R.A., & Just,

- M.A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, 320, 1191-1195.
- Murphy B., Baroni M. & Poesio M. (2009). EEG responds to conceptual stimuli and corpus semantics. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)* (pp. 619-627). East Stroudsburg, PA: Association for Computational Linguistics.
- Nolan, J.R.M. (1958). *An examination of basic Freudian principles in the light of Thomistic psychology*. M.A. dissertation, University College, Dublin.
- O'Doherty, E.F. (1962). Psychoanalysis, psychotherapy and spiritual direction. In E.F. O'Doherty & S.D. McGrath (Eds.), *The priest and mental health* (pp. 66-86). Dublin: Clonmore and Reynolds.
- Palace, E.M. & Gorzalka, B.B. (1990). The enhancing effects of anxiety on arousal in sexually dysfunctional and functional women. *Journal of Abnormal Psychology*, 99, 403-411.
- Plé, A. (1952). Saint Thomas Aquinas and the psychology of Freud. *Dominican Studies*, 5, 1-34.
- Romppel, M. (2001). *CoAn für Windows 2.07*. Bielefeld: Matthias Romppel.
- Scandella, L. (2004). Les cycles longs et les œuvres de pensée. *Cahiers du CEP*, 10, 74-97.
- Skitka, L.J. & Sargis, E.G. (2005). Social psychological research and the internet: The promise and peril of a new methodological frontier. In Y. Amichai-Hamburger (Ed.), *The social net: The social psychology of the internet* (pp. 1-25). New York: Cambridge University Press.
- Stolorow, R.D. (1975). Addendum to a partial analysis of a perversion involving bugs: An illustration of the narcissistic function of perverse activity. *International Journal of Psycho-Analysis*, 56, 361-364.
- Turnbull, O.H. & Solms, M. (2007). Awareness, desire, and false beliefs: Freud in the light of modern neuropsychology. *Cortex*, 43, 1083-1090.
- Van de Vijver, G. (2000). About psychoanalysis, thunderstorms, and other pleasurable things: An interview with Karl Pribram, August 13th, 1999. *Psychoanalytische Perspectieven*, 39, 91-112.
- Wedemann, R.S., de Carvalho, L.A.V. & Donangelo, R. (2008). Network properties of a model for conscious and unconscious mental processes. *Neurocomputing*, 71, 3367-3371.
- Westen, D. (1998). The scientific legacy of Sigmund Freud: Toward a psychodynamically informed psychological science. *Psychological Bulletin*, 124, 333-371.
- Whitelaw, J.D.A. (1959). A case of fetishism treated with lysergic acid diethylamide. *Journal of Nervous and Mental Disease*, 129, 573-577.
- Wilson, A. (2009). *Corpus evidence for aspects of the theory of fetishism*. Unpublished manuscript.
- Wise, T.N. (1985). Fetishism - etiology and treatment: A review from multiple perspectives. *Comprehensive Psychiatry*, 26, 249-257.
- Wolters, M. & Kirsten, M. (1999). Exploring the use of linguistic features in domain and genre classification. In *Proceedings of EACL 99*, pp. 142-149.
- Wulff, M. (1946). Fetishism and object choice in early childhood. *Psychoanalytic Quarterly*, 15, 450-471.