Influence of the field of education on meaning predictability of novel compounds

Lenka Janovcová

This paper discusses the impact of language users' field of education on their meaning prediction processes. It's main aim is to compare students of natural sciences with students of humanities in terms of meaning predictability (Štekauer, 2005). The results of the experiment, based on an analysis of a sample of ten English non-existing novel compounds, indicate that the field of education of language users belongs to extra-linguistic factors influencing meaning prediction of novel naming units. Differences in meaning prediction between the studied groups observed on the experimental sample of new compounds are both of quantitative and qualitative nature.

Keywords: *meaning predictability, novel compounds, natural sciences, humanities*

1. Introduction

When language users encounter a new lexical unit for the first time they assign certain lexical meaning to it. Since the lexical unit is unknown to them, such an assignment of meaning is a matter of prediction. However, the meaning is not assigned arbitrarily – apart from the grammar rules of a language, language users' prediction is governed by their previous experience. In his meaning predictability theory, Štekauer (2005) emphasizes the role of extra-linguistic knowledge, experience and personal preferences in the process of meaning interpretation. Language users judge plausibility of certain interpretations in accordance with their extra-linguistic knowledge and experience. These judgments are directly influenced by the reality experienced by interpreters – plausible are those readings which reflect more realistic relations than others and denote possible extra-linguistic objects.

Extra-linguistic knowledge of language users depends on their position in society. Gleitman and Gleitman (1970) found out that educational level of language users has impact on their interpretation of new compounds (lower educated participants of their research made more 'errors' – proposals of unacceptable readings – than participants with higher education). These results indicate that (at least some) sociolinguistic factors belong to extra-linguistic factors that influence the interpretation process. Other relevant sociolinguistic factors include age, occupation, mother tongue (especially in the case of non-native speakers), etc. Štekauer (2005) also demonstrates that the native/non-native speaker factor does not play any significant role in interpretation of novel English naming units. This study focuses on the language users' field of education, in particular, compares the field of natural sciences and the field of humanities.

It is reasonable to expect that this factor may influence interpretation of novel naming units. Karakas (2010) points out that over the last fifty years there has been a discussion on differences between sciences and humanities in terms of the paradigms of thinking. He points out that traditionally, *creative thinking* has been associated with humanities, while *critical thinking* with sciences. Karakas as well as Baker and Rudd (2001) give an overview of characteristics assigned to creative and critical thinking. Critical thinking is reflective and analytical, based on careful consideration of supposed knowledge and logical reasoning. Its aim is to achieve a certain goal – to decide what to think or do. Although critical thinkers

should be open to alternatives, they concentrate on finding a solution ('a desirable outcome'), and thus critical thinking is considered convergent. On the other hand, creative thinking is inventive, imaginative, intuitive, based on originality and a sudden 'illumination' which follows an unconscious phase of incubation (unconscious thought of a problem). Creative thinking tries to provide various possibilities and original ideas and therefore it is considered divergent. Nevertheless, in the process of creative problem solving, the phase of creativity is always preceded and followed by critical, analytic thinking (in the form of initial study of data and sources, and subsequent verification of hypotheses). Although Karakas (2010) argues that people from science as well as humanities use both creative and critical thinking and that these processes of thinking overlap, Mumford et al. (2010) found some differences in several stages of thinking during creative problem solving process among students of different fields.

Mumford et al. studied thinking processes in doctoral students of health, biological and social sciences. While students of the first two categories might be considered to belong to the field of sciences (although the authors considered them distinct fields, as health science is applied), students of social sciences belong to the field of humanities. The researchers tested several abilities needed for individual phases of problem solving (i.e., problem definition, information gathering, information organization, conceptual combination, idea generation, idea evaluation, implementation planning, solution appraisal). Out of these variables, conceptual combination is the most important for meaning predictability of new primary compounds.

This paper studies meaning predictability of new lexical units on a sample of non-existing English noun-noun compounds. Primary compounds are relevant when studying extra-linguistic factors of the meaning-prediction process because knowledge of the meaning of their constituent parts is not sufficient for their understanding. To interpret a compound, it is important to identify a certain relationship between the compound stems. Lieber claims that it is generally agreed by linguists that "there is no fixed semantic relationship between the two stems of a root compound" (2005: 380). Thus, when a language user (listener) encounters a compound for the first time, he/she has to choose one of its possible interpretations by putting the stems into a certain semantic relation. From the psycholinguistic point of view (c.f. Gagné and Shoben 1997; Gagné and Spalding 2006), primary compounds are combined concepts. That said, interpretation of a compound is a result of conceptual combination that includes the selection of the thematic relation which links the compound constituents

In Mumford et al. (2010), conceptual combination is assessed through the quality and originality of new concepts produced by a combination of concepts in the problem solution, e.g. birds and sporting equipment (2010: 15). It was found out that biological and health science students obtained lower scores on conceptual combination than social science students (ibid.: 21). These results suggest that there might be some differences in thinking that distinguish sciences students from humanities students.

The experiment described in this paper is aimed at the comparison of humanities students and natural sciences students in terms of meaning-prediction processes (on a sample of context-free novel compounds). It aims to show that the field of education of language users belongs to those extra-linguistic factors that influence meaning prediction of novel naming units. Apart from comparing the studied groups, the research aims to reveal general tendencies of interpretation of novel English noun-noun compounds.

2. The Experimental Paradigm

The experimental sample is composed of ten novel English compounds created by the author for the sake of this experiment, i.e. *hand tomato, hair computer, table shoes, chair bread, car egg, apple leg, street potato, sky piano, horse tree, dog paper*. Their non-existence was tested against Macmillan English Dictionary for Advanced Learners (2007) and the on-line version of British National Corpus (http://www.natcorp.ox.ac.uk/) and they were studied outside any context. Studying predictability of novel naming units without a context may provide insights into the meaning predictability process in general – Štekauer (2005) claims that the naming act is of cognitive nature and the core, denotative meaning a newly created lexical unit is stable in every context. Context-free naming units allow generalisations as they are 'pure' results of the naming process, while context may lead to unique, idiosyncratic meanings heavily dependent on the context itself.

Seventy-two undergraduate students of Pavol Jozef Šafárik University in Košice, Slovakia, participated in the experiment. The experiment was carried out during seminar classes. It took about 25 minutes, i.e. the time during which the majority of the participants were able to complete their task. All the participants were non-native speakers of English with Slovak as their first language. Although their level of proficiency in English was not pre-experimentally measured, only questionnaires meaningfully filled in English were accepted. This criterion was taken as a guarantee of sufficient knowledge of English for the purpose of the experiment. Štekauer (2005) shows that when non-native language users from the same socio-cultural background as native speakers (Western culture) have 'standard command' of English (a non-native speaker understands the meaning of morphemes and he/she is aware of the word-formation rules (ibid.: 61)), the level of their individual proficiency does not influence meaning predictability. All compounds were created by combining nouns that belong to the core vocabulary of English. Instructions were given in both English and Slovak. Participants were also provided oral instructions and were informed that their responses would be treated confidentially.

After elimination of the questionnaires that were filled in Slovak and those which were not filled meaningfully, the questionnaires of 60 informants qualified to be used in the experiment. Thirty of these informants were students of English at the Faculty of Arts. Some of them were single majors in English and some of them were double majors studying English in combination with a subject belonging to Humanities (another language, philosophy, psychology). This group was labelled 'Humanities'. A group of the other thirty undergraduates studied mathematics at the Faculty of Sciences. Some of them studied only mathematics, while others studied mathematics in combination with some other subject belonging to natural sciences (information science, biology, physics, geography, chemistry). This group of participants was labelled 'Science'. Science and Humanities together as one group are here referred to as 'Total'.

Informants were asked to write down (in English) as many possible meanings of each novel compound word as they could think of. For each of the proposed meaning, they were asked to assign a score from 10 points (indicating very high probability of occurrence of such a meaning in the language), to 1 point (indicating minimum chance of occurrence of such a meaning in the language). The results were evaluated within the framework of meaning predictability theory (Štekauer 2005). For the purpose of comparing meaning predictability of

different readings (interpretations) of new naming units, Štekauer designed tools called the Predictability Rate (PR) and the Objectified Predictability Rate (OPR).

Computation of the Predictability Rate serves for comparison of different readings belonging to the same novel naming unit. Its calculation is based on two postulates: "The predictability of meanings of naming units correlates with the acceptability of these meanings to interpreters." and "Since there is no clear-cut boundary between acceptable and unacceptable meanings the predictability of the meanings of naming units is a *cline*." Predictability Rate is then calculated according to the following formula:

(1) $$PR = \frac{r}{R_{max}} \times \frac{p}{P_{max}}$$

where:
r = the number of informants identifying a particular meaning as acceptable
R_{max} = the total number of informants
p = the sum total of the points assigned to a given meaning by all informants (on a scale from 1 to 10, where 10 stands for the highest acceptability of the meaning)
P_{max} = the maximum possible number of points assignable by all informants (Štekauer 2005: 94-95).

The value of $\frac{r}{R_{max}}$ represents the frequency of occurrence of a given interpretation. For illustration, let's take a novel compound *hand tomato*. Out of 60 participants, 23 suggested a reading 'a tomato of a size of a hand (as big as a hand)'. Therefore for this interpretation, *r* = 23; R_{max} = 60 and the frequency of occurrence is 23/60 = 0.38. The total sum of points assigned to this meaning was 128, while the maximum possible number of points was 10x60 = 600. Thus p = 128 and P_{max} = 600. When these values are instituted in the formula, the value of PR is calculated as follows: PR = (23/60) x (128/600) = 0.082.

While PR is a tool for comparison of various readings of the same novel naming unit, the Objectified Predictability Rate serves for comparison of the most predictable readings belonging to various naming units. The underlying idea of OPR is based on the theory of competition, i.e. various readings of a naming unit compete among themselves for the dominant position. The tougher the competition is (the more interpretations have relatively high PR), the less clear it is which reading is the dominant one. By implication, OPR expresses the 'level of competition' of various readings of a lexical unit, and its value for a given lexical unit is calculated as a ratio of the PR^{top} (PR of the reading with the highest value of PR), and the sum of PR^{top}, PR^{top-1}, PR^{top-2} (PRs of the three readings with the highest values of PR). The formula for calculating OPR is as follows:

(2) $$OPR = \frac{PR^{top}}{PR^{top} + PR^{top-1} + PR^{top-2}}$$

By means of OPR, it is possible to compare the top readings of different naming units - the higher the OPR the lower is the competition between the top readings and the higher is the actual predictability of the most acceptable reading. From this it follows that a high value of PR^{top} does not guarantee a high value of OPR – if there are several readings with a high PR, the OPR is relatively low. On the other hand, if PR^{top} is relatively low in a naming unit but PRs of all other readings are far lower, the value of OPR is comparably high. Thus, OPR

makes it possible to meaningfully compare naming units with readings that achieve significantly distinct absolute predictability rates. Consequently, the use of OPR might help to reveal a more complex picture of meaning predictability in general than the use of PR only.

Moreover, Štekauer emphasizes the significance of the gap between the absolute values of PRs for different readings of the same naming unit, and labels it as the Predictability Rate Gap (PRG). For example, PRG between the two most predictable readings equals PR^{top} - PR^{top-1}. Similarly to OPR, calculation of PRG between the most predictable reading and subsequent readings indicates dominance of the most predictable reading (if PRG between the most predictable and the second most predictable reading is significant) or competition between the top readings (if the mentioned PRG is low).

Last but not least, Štekauer's meaning predictability theory hypothesises that meaning predictability of novel compounds is a cline with one or two central (typical and hence highly predictable) meanings for each novel compound that are preferred by a relatively homogeneous group of language users (a group of language users with similar level and field of education, of similar age living in a certain cultural environment – ibid.: 98). These central meanings may or may not be the same for the two tested groups (Humanities students and Science students). However, they should be the same within the group. Humanities students often incline to more creative and divergent thinking than Sciences students, while science students tend to concentrate on achieving the aim using critical thinking (c.f. Karakas 2010). Therefore it was expected that there are quantitative differences between the results of Humanities and Sciences students – Humanities students were expected to propose a higher average number of possible meanings for individual novel compounds.

3. Results and discussion

3.1 *Statistics*

The results, summarized in Table 1, point to both quantitative and qualitative differences in meaning predictability within the Sciences and the Humanities groups. Readings assigned to individual sample compounds most frequently, together with some of the respective statistical indicators, can be found in the Appendix. Table 1 provides mean values of studied mathematical indicators and thus opens space for observation of general tendencies.

Line 1 of Table 1 gives the number of different readings proposed by informants within a given group. Lines 2 and 3 show the proportion of common readings proposed both within Sciences and Humanities. Lines 4 and 5 deal with the proportion of idiosyncratic and non-idiosyncratic readings of individual compounds. 'Non-idiosyncratic' readings are those assigned to a novel compound by at least 3 informants belonging to the same group (Science or Humanities), all other readings are referred to as 'idiosyncratic'. The PR values of idiosyncratic readings approximate zero (they rarely exceed the value 0.01) thus they do not reveal general tendencies in meaning prediction of novel compounds. Many of idiosyncratic readings were proposed by the only informant (as indicated in line 6).

Line 7 gives an average number of proposed readings per informant within studied groups while lines 8 – 10 show Predictability Rates of the readings with highest values of this marker. A top reading is a reading with the highest value of PR (PR^{top}). Line 11 indicates whether the groups of Sciences and Humanities proposed the same top reading. On the other

hand, line 12 gives information whether the top reading of a compound is dominant or it competes with other readings. If there is a dominant reading in Sciences, it is labelled D1. If Humanities or (and) Total has the same dominant reading, a corresponding field is filled with D1 as well. Similarly, if there is a dominant reading for Humanities, it is labelled D2 (an exception is the case when D2 is the same as the dominant reading of Sciences – in such a case it is labelled D1). If in Total there is the same dominant reading as in Humanities, it is also labelled D2. If the dominant reading for a certain studied group can not be identified, the corresponding field is labelled C for 'competition among several readings'.

An indicator of the degree of competition among the most predictable readings is the value of OPR (line 13). Of course, competition of readings is a scale from basically no competition to 'tough competition' of readings with almost equal PRs. For clearly competing readings, OPR around 0.3 – 0.4 was observed, while for the dominant readings with minimal competition around 0.7 – 0.9. For simplicity, if the most predictable reading gained OPR higher than 0.5, it is considered a dominant reading; if its OPR is lower than 0.5, it is considered one of the competing readings (the values around 0.5 were detected for the compounds where PR^{top-1} is about one half of PR^{top}, and such a difference between the readings can be interpreted as the dominance of the most predictable reading, although the absolute Predictability Rate Gap might be low).

To decide whether the most predictable reading is dominant, Štekauer (2005) uses not only the value of OPR, but also the ratios of PR^{top}/PR^{top-1} and PR^{top}/PR^{top-2}. Although in the sample studied in this thesis the values of PRs are generally very low, the ratios of PR^{top}/PR^{top-1} and PR^{top}/PR^{top-2} support the criterion defining a dominant reading as an interpretation with OPR higher than 0.5 (with the only exception of *horse tree* – its PR^{top} is only 0.496, but it can be rounded to 0.5). For the dominant reading, the first ratio is always higher than 1.28 (but mostly it is much higher) and at the same time, the second ratio is more than 2.4, while for competing readings, both ratios are lower than 2 (lines 14 – 16). The average values of these ratios are high, because some readings are as dominant as having the PR^{top} more than ten times higher than PR^{top-1}. However, there is a great difference when the average values of dominant readings are compared with those of competing readings.

Based on the data presented in the table, several tendencies can be observed. First of all, it is important to notice that all the tested novel compounds belong to unpredictability level, defined by Štekauer (2005) as the level of lexical units with PR^{top} occurring in the interval 0 – 0.25[1]. The highest PR in the studied sample is 0.135, the second highest 0.116 and the third 0.1 – all of the rest PRs are lower than 0.1. On the basis of his data, Štekauer (ibid., p. 13) concluded that "the *general tendency for the PR of primary compounds*, … , *is the PR value about 0.3"*. However, the average PR^{top} of the sample in the present study is in Total (looking at Science and Humanities together) only 0.034 (Table 1, line 10). Thus, it is probable that the tendency for the PR values depends on the level of predictability to which studied nominal compounds pertain.

[1] Štekauer (2005: 153) divides lexical units into four 'predictability levels' according to the value of their PR^{top}: unpredictability level (0 - 0.25); low predictability level (0.26 –0.5); medium predictability level (0.51 – 0.75); high predictability level (0.76 – 1).

			hand tomato	hair comp.	table shoes	chair bread	car egg	apple leg	street potato	sky piano	horse tree	dog paper	Average
1	Number of proposed readings	Sc./ Hum./ Tot.	21 36 47	15 33 34	14 34 40	21 32 42	16 34 42	23 39 52	20 29 40	19 37 43	19 34 43	24 40 48	19,2 34,8 43,1
2	Common Sc. & Hum.		10	14	8	11	8	10	9	13	10	16	10,9
3	% of common readings		21,28	41,18	20	26,19	19,05	19,23	22,5	30,23	23,26	33,33	25,62
4	number of non-idiosyncratic readings		8	12	11	9	7	8	7	6	4	9	8,1
5	% of idiosyncratic r.		82,98	64,71	72,5	78,57	83,33	84,62	82,5	86,05	90,7	81,25	80,72
6	% of 1-informant r.		59,6	35,3	65	64,3	66,7	69,2	72,5	60,5	69,7	58,3	62,11
7	Number of readings per informant	Sc./ Hum./ Tot.	1,3 2,5 1,9	1,2 2,33 1,765	1,1 2,4 1,75	1 1,8 1,4	1,1 2,1 1,6	1 2,13 1,565	1,3 2,43 1,865	1 2,03 1,515	0,9 2 1,45	1,2 2,27 1,735	1,11 2,2 1,65
8	PR^{top} PR^{top-1} PR^{top-2}	Sc.	0,041 0,008 0,006	0,039 0,009 0,006	0,018 0,016 0,015	0,004 0,004 0,004	0,085 0,008 0,001	0,012 0,005 0,003	0,02 0,016 0,011	0,029 0,004 0,003	0,013 0,005 0.0007	0,027 0,013 0,005	0,028 0,010 0,005
9	PR^{top} PR^{top-1} PR^{top-2}	Hum.	0,135 0,012 0,01	0,028 0,025 0,02	0,045 0,02 0,018	0,024 0,01 0,006	0,1 0,013 0,008	0,02 0,013 0,013	0,116 0,062 0,048	0,04 0,026 0,008	0,059 0,046 0,014	0,023 0,019 0,019	0,059 0,025 0,016
10	PR^{top} PR^{top-1} PR^{top-2}	Tot.	0,082 0,009 0,007	0,021 0,011 0,007	0,029 0,017 0,01	0,006 0,006 0,005	0,093 0,005 0,004	0,013 0,01 0,009	0,054 0,029 0,029	0,016 0,012 0,011	0,005 0,003 0,003	0,023 0,008 0,007	0,034 0,011 0,009
11	Same top-reading for Sc. & Hum.		yes	no	no	no	yes	no	no	no	no	no	
12	Dominant reading vs. competition	Sc./ Hum./ Tot.	D1 D1 D1	D1 C D1	C D2 D2	C D2 C	D1 D1 D1	D1 C C	C D2 C	D1 D2 C	D1 D2 C	D1 C D1	
13	OPR	Sc./ Hum./ Tot.	0,747 0,857 0,836	0,718 0,379 0,54	0,367 0,536 0,518	0,355 0,58 0,36	0,905 0,822 0,912	0,612 0,433 0,416	0,44 0,513 0,482	0,811 0,544 0,404	0,714 0,496 0,471	0,604 0,378 0,609	0,627 0,554 0,555
14	PR^{top}/PR^{top-1} PR^{top}/PR^{top-2}	Sc.	5.325 6.613	4.333 6.5	1.125 1.2	1.05 1.135	10.63 85	2.48 4.276	1.282 1.887	7.175 10.63	2.66 19	2.116 5.46	3.8171 14.17
15	PR^{top}/PR^{top-1} PR^{top}/PR^{top-2}	Hum.	11.25 13.5	1.12 1.4	2.25 2.5	2.4 4	7.692 12.5	1.504 1.55	1.871 2.417	1.545 5.224	1.283 4.245	1.201 1.233	3.2115 4.8568
16	PR^{top}/PR^{top-1} PR^{top}/PR^{top-2}	Tot.	9.111 11.71	1.909 3	1.706 2.9	1 1.2	18.6 23.25	1.3 1.444	1.862 1.862	1.333 1.455	1.852 1.923	2.875 3.286	4.1548 5.2034

Table 1 *Statistics*

Although average PR values of the studied data do not approach values based on Štekauer's (2005) research, Štekauer's hypothesis saying that there is a tendency to have one central (highly predictable) meaning for each novel compound preferred by a relatively homogeneous group of language users, is supported by the obtained data. Sciences informants proposed a dominant reading for seven compounds out of ten (Table 1, Line 12),

Humanities participants also for seven. Although these compounds overlap only partially, there is a clear tendency for one dominant reading within each group.

In Total, there are only 5 dominant readings. However, this finding does not contradict the general tendency of one dominant reading within a homogeneous group of language users, as Total group is not considered homogeneous due to distinct fields of education of its members. A relatively low proportion of dominant readings in Total is caused by the fact, that with the exception of *hand tomato* and *car egg*, all novel compounds have either a dominant reading in one studied group (Humanities or Sciences) and competition in the other (*hair computer, table shoes, chair bread, apple leg, street potato*), or there are two different dominant readings for Science and Humanities (*sky piano, horse tree*). Increased competition between the readings is influenced also by the unpredictability level of studied compounds which contributes to the fact that the informants of none of the experimental groups were able to give a dominant interpretation for three out of ten readings. In such cases, there are mostly three mutually competing readings (for *street potato* in Sciences, there are only two). In Total, there is the same tendency – if there is no dominant reading, there are mostly three readings that compete (in the case of *hand tomato*, there are potentially four competing readings but when computing the OPR only the first three readings are taken into account; for *horse tree* there are only two competing interpretations).

The hypothesis that central meanings may or may not be the same for the two tested groups (Humanities and Sciences) is supported by the obtained results as well. There are only two compounds sharing the same most predictable reading – *hand tomato* and *car egg* (Table 1, Line 12). In several cases, the most predictable reading within one group of informants is not even proposed (or its PR is extremely low) in the other group. In six cases, there is a dominant reading proposed by one group while there is a competition among the readings proposed by the other group. In two cases, Sciences and Humanities proposed different dominant readings, and only in two cases both groups proposed the same dominant reading. These observations point to qualitative differences between the Sciences and Humanities groups in terms of their meaning prediction abilities.

Another hypothesis, that there are quantitative differences between the results of the Humanities and Sciences students, is acknowledged by the experimental data as well. The Humanities students not only proposed a higher average number of possible meanings for individual novel compounds, but they also propose considerably more different interpretations of novel compounds per informant - an average number of readings proposed by one informant in Sciences is 1.11 (Table 1, Line 7). For Humanities it is twice as many: 2.2. The Sciences informants proposed on average 19.2 different readings for each novel compound, while the Humanities participants proposed almost twice as many – in average 34.8 readings per compound. In Total, an average number of readings proposed for each novel compound is 43.1. However, in average only 25.62% of these readings overlap in the groups of Sciences and Humanities (Table 1, Lines 1, 3) – this proportion correlates with a high proportion of idiosyncratic readings. The observation that the Humanities informants proposed twice as many different readings for a compound (in terms of the overall number of interpretations as well as of the number of readings per informant) gives support to the assumption that there are some differences of thinking between students of Sciences and Humanities (c.f. Mumford et al. 2010). It seems that creative thinking of the Humanities informants is divergent, aimed at seeking several possible options, while the Sciences students usually concentrate on finding one solution.

Another quantitative difference between the studied groups concerns Predictability Rates. The average value of PR^{top} is twice higher for Humanities than for Sciences (0.59 : 0.29), the average value of PR^{top-1} is 2.77 times higher (0.025 : 0.009), and the average value of PR^{top-2} is higher 3.2 times (0.016 : 0.005) (Table 1, Lines 8, 9). These values suggest that the Humanities informants find readings they propose more probable. The mean values of AI and NA support this observation - the average number of points given by one informant (AI) for all non-idiosyncratic readings of all compounds is 3.63 for Sciences, and 5.69 for Humanities. The average number of informants accepting the same reading (NA) for all non-idiosyncratic readings of all compounds is 2.17 for Sciences and 4.35 for Humanities. Moreover, all three readings with PR higher than 0.1, appear among Humanities PRs.

As shown by Štekauer (2005), the PR and OPR values of any reading are not proportional. In the obtained data, the average value of the OPR is slightly higher for Sciences (0.627) than for Humanities (0.554). These values suggest that in Humanities, the competition among the readings is much tougher. This assumption is also supported by the ratios of PR^{top}: PR^{top-1}: PR^{top-2} within Sciences and Humanities. These ratios are 5.8 : 1.8 : 1 for Sciences, and 3.7 : 1.6 : 1 for Humanities. Although the ratios between the second and the third readings are comparable, the ratio between the first and the second readings is much higher for Sciences which suggest that the position of the most predictable reading in Sciences is generally 'more dominant' than in Humanities.

3.2. *Unpredictability level*

Apart from the above-mentioned statistical observations, the unpredictability level of the studied compounds is connected with several findings about the tested compounds. Firstly, all studied compounds have a very high proportion of idiosyncratic readings. An average proportion of idiosyncratic readings is 80.72% (Table 1, Line 5) - the highest proportion is as high as 90.7% (*horse tree*). A vast majority of readings is idiosyncratic as a consequence of language users' difficulties with identifying an applicable relation between the compound stems. Thus, many idiosyncratic readings are unique and original, very specific for the way of thinking and world experience of their authors.

Secondly, the unpredictability level of the studied compounds is reflected in the distribution of the points assigned to individual readings. The points are relatively evenly dispersed from low to high values – they are not concentrated around a constant value. Dispersion of the assigned points suggests that informants do not perceive the 'goodness' of a given reading in a similar way, and it is not possible to identify any semantic relation between the nouns expressed in a reading on a typicality scale. This observation is quantitatively expressed by standard deviation (SD) – see Appendix. The values of standard deviation are high for both experimental groups – its average value is 1.57 for Sciences and 2.18 for Humanities. Even though several informants came up with the same interpretation, they were not certain about the level of its acceptability (probability of occurrence). This fact may support the assumption that in the case of unpredictability level compounds, interpretation is not narrowed down by a prototypical (easily accessible) relation.

Thirdly, the unavailability of a typical relation might be connected not only to dispersion of assigned points but also to the presence of many metaphorical readings in the obtained data. It is probable that when informants could not easily find a literal representation for a compound in extra-linguistic reality, they created meanings based on associations. These associations are rooted in their extra-linguistic knowledge or created on the basis of an

established compound which could serve as a source for analogy. This approach to compound processing can account for many figurative meanings, a high proportion of exocentric interpretations of *street potato* and *apple leg* (analogy with *couch potato* or *skinhead, redhead*), as well as for some coordinative interpretations (such as 'a tree with a horse hanging on it (a toy)' for *horse tree*, or 'an egg on wheels' for *car egg*) present it the obtained data. English primary compounds are in general right-headed and thus they are usually endocentric. Despite this fact, there are some exocentric (*read head*) and copulative (*producer-director*) root compounds in English (Lieber 2005: 375-379). Nevertheless, if language users could not find any available relation between the compound stems, they did not give any reading or they proposed an unacceptable interpretation (mostly by making the left-hand constituent the head of a compound).

Although there are many figurative interpretations, the relation between the compound constituents of the majority readings reflects a referent that might exist in extra-linguistic reality. There is a strong tendency for compositionality of the most predictable readings. All of them, except for *car egg*, are clearly acceptable. *Car egg*'s most predictable reading ('a car that looks like an egg') can be considered acceptable only in its coordinative sense – 'a car which is an egg (in the sense of the shape)'. This explanation is probable because other readings show that the informants were aware of the general rules underlying interpretation of compounds (English nominal compounds are in general endocentric with the right-hand stem being the head of a compound). Despite a strong tendency for compositionality of the most predictable readings, the unpredictability level of compounds causes that it is difficult to match compound stems transparently. Therefore some of the most predictable interpretations are figurative ('slow breakfast' for *chair bread* in Sciences, 'a disease causing a leg to look like an apple' for *apple leg* by Humanities, 'a beautiful piano song' for *sky piano* by Humanities) and, one reading is based on analogy ('a person wandering in the streets' for *street potato* – based on *couch potato*).

The fact that besides compositional readings there are numerous figurative readings indicates that the unpredictability level can be considered to be a good basis for studying the impact of language users' extra-linguistic knowledge on meaning predictability. Of course, language users need extra-linguistic knowledge to choose a plausible relation between the nouns in a compositional reading. However, when creating a figurative reading, their associations (based on common or individual extra-linguistic experiences) can be traced (in most of the cases). In such cases, extra-linguistic experience does not influence only the choice of an acceptable relation between the nouns, but also the preference for associations connected to the concepts of these nouns. For example, one of the Sciences informants proposed an exocentric interpretation for four of ten studied compounds, which is a very high proportion comparing to other informants. The source of such interpretations can not be looked for in the knowledge of language (word-formation rules) but in his/her experience. The figurative interpretations of novel compounds observed in the studied data indicate that Libben's (1998) hypothesis, that all novel lexical units are transparent, seems not to be universal. It is valid in the majority of the studied cases, but not in all.

4. Conclusion

The research suggests that the field of education of language users belongs to extra-linguistic factors that influence the meaning prediction of novel naming units belonging to the

unpredictability level. In the studied sample of novel compounds, meaning predictability of Sciences and Humanities students differ in both quantitative and qualitative terms. However, further research is needed to confirm these findings. For example, the results of a similar experiment might be sensitive to the field of education in research that will deal with compounds divided according to their level of interpretability (predictability). It is possible that if a relation between the stems of nominal compounds (based on their prototypical features) is readily accessible, the difference in the number of proposed readings among Humanities and Sciences would not be as significant as in the case of unpredictability level compounds. Nevertheless, the research shows that the field of education is a factor relevant to the meaning-prediction process.

References

BAKER, Matt, RUDD, Rick. 2001. "Relationships between Critical and Creative Thinking." In *Journal of Southern Agricultural Education Research 5 (1)*, 2001, pp. 173-188.

British Natinal Corpus. Available online < http://www.natcorp.ox.ac.uk/> [Cit. 5.11. 2012].

GAGNÉ, Christina. L., SHOBEN, Edward J. 1997. "Influence of Thematic Relations on the Comprehension of Modifier-Noun Combinations". In *Journal of Experimental Psychology: Learning, Memory, and Cognition 23(1)*, 1997, pp. 71-87.

GAGNÉ, Christina L., SPALDING, Thomas L. 2006. "Conceptual Combination: Implications for the mental lexicon." In Libben, G. – Jarema, G. (eds.) *The Presentation and Processing of Compound Words*. New York: Oxford University Press, 2006a, pp. 145-168.

GLEITMAN, Lila R., GLEITMAN, Henry. 1970. *Phrase and Paraphrase: Some Innovative Uses of Language*. New York: W. W. Norton and Co., 1970.

KARAKAS, Scott L. 2010. "Creative and Critical Thinking in the Arts and Sciences: Some Examples of Congruence." *Forum on Public Policy Online* 2010 (2). Available online <http://forumonpublicpolicy.com/spring2010.vol2010/spring2010archive/karakas.pdf> [Cit. 5.11. 2012].

LIEBER, Rochelle. 2005 "English word-formation processes." In Štekauer, P. – Lieber, R. (eds.) *Handbook of Word-Formation*. Dordrecht: Springer, 2005.

LIBBEN, Gary. 1998. "Semantic Transparency in the Processing of Compounds: Consequence for Representation, Processing and Impairment." In *Brain and Language 61*, 1998, pp. 30-44.

Macmillan English Dictionary for Advanced Learners. Second edition. Oxford: Macmillan Education, 2007.

MUMFORD, Michael D., ANTES, Alison L., CAUGHRON, Jared J., CONNELLY, Shane, BEELER, Cheryl. 2010. "Cross-Field Differences in Creative Problem-Solving Skills: A Comparison of Health, Biological, and Social Sciences." In *Creativity Research Journal 22(1)*, 2010, pp.14-26.

ŠTEKAUER, Pavol. 2005. *Meaning Predictability in Word-formation*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 2005.

Appendix
Proposed readings

Readings – all non-idiosyncratic readings of the studied compounds
OPR – Objectified Predictability Rate of individual naming units
SUM – the sum of all points assigned to a reading by all 30 informants within the group
NA – the number of informants accepting the meaning (the number of informants that proposed a given reading)
AI – average number of points given by one informant (AI = SUM/NA)
SD – standard deviation
PR – Predictability Rate of individual readings

Readings	**Science**					**Humanities**					**Total**			
hand tomato	SUM	NA	AI	SD	**PR**	SUM	NA	AI	SD	**PR**	SUM	NA	AI	**PR**
a tomato of a size of a hand (as big as a hand)	41	9	4,6	2	**0,041**	87	14	6,2	2,5	**0,135**	128	23	5,6	**0,082**
ketchup (a tomato pressed by a hand)	23	3	7,7	2,5	**0,008**	24	4	6	2,8	**0,011**	47	7	6,7	**0,009**
a tomato that people throw on others	14	4	3,5	3,3	**0,006**	17	4	4,3	2,5	**0,008**	31	8	3,9	**0,007**
a ball used in a new game	14	2	7	0	**0,003**	22	4	5,5	2,1	**0,01**	36	6	6	**0,006**
a red hand	4	2	2	1,4	**0.0009**	24	4	6	1,8	**0,011**	28	6	4,7	**0,005**
a tomato in a shape of a hand	4	1	4		**0.0004**	16	4	4	1,4	**0,007**	20	5	4	**0,003**
a small tomato	0	0	0		**0**	27	4	6,8	3,3	**0,012**	27	4	6,8	**0,003**
a device for cutting tomatoes	0	0	0		**0**	20	3	6,7	0,6	**0,007**	20	3	6,7	**0,002**
OPR	**0,747**					**0,857**					**0,836**			

hair computer	SUM	NA	AI	SD	**PR**	SUM	NA	AI	SD	**PR**	SUM	NA	AI	**PR**
a computer that generates the best hair-style for somebody	44	8	5,5	2,4	**0,039**	24	3	8	1,7	**0,008**	68	11	6,2	**0,021**
a very thin computer	0	0	0		**0**	36	7	5,1	1,8	**0,028**	36	7	5,1	**0,007**
a mechanical hair-dresser	0	0	0		**0**	38	6	6,3	2,3	**0,025**	38	6	6,3	**0,006**
a computer that determines the quality of one's hair	11	3	3,7	1,5	**0,004**	37	5	7,4	2,6	**0,021**	48	8	6	**0,011**
a miniature computer worn on hair	0	0	0		**0**	29	4	7,3	1,5	**0,013**	29	4	7,3	**0,003**
a computer that presents many hair-styles	7	2	3,5	2,1	**0,002**	25	4	6,3	3,8	**0,011**	32	6	5,3	**0,005**
a machine that dries and straightens hair	21	4	5,3	1,5	**0,009**	6	1	6		**0.0007**	27	5	5,4	**0,004**
a computer connected to other computers by long cables	0	0	0		**0**	25	3	8,3	1,5	**0,008**	25	3	8,3	**0,002**
an instrument that makes one's hair look beautiful	0	0	0		**0**	17	3	5,7	2,5	**0,006**	17	3	5,7	**0,001**
an old computer	0	0	0		**0**	14	3	4,7	3,8	**0,005**	14	3	4,7	**0,001**
a computer that counts the number of hair on one's head	18	3	6	3,6	**0,006**	12	2	6	4,2	**0,003**	30	5	6	**0,004**
a computer made of hair	11	3	3,7	2,3	**0,004**	4	2	2	0	**0.0009**	15	5	3	**0,002**
OPR	**0,718**					**0,379**					**0,54**			

	Science					Humanities					Total			
table shoes	SUM	NA	AI	SD	**PR**	SUM	NA	AI	SD	**PR**	SUM	NA	AI	**PR**
shoes for the table (to prevent the table from shaking)	30	5	6	1,2	**0,017**	45	9	5	2,9	**0,045**	75	14	5,4	**0,029**
shoes designed to be worn on special occasions (like dinners)	0	0	0		**0**	31	6	5,2	3,3	**0,021**	31	6	5,2	**0,005**
shoes that one wears when one wants to put their feet on the table	0	0	0		**0**	33	5	6,6	1,1	**0,018**	33	5	6,6	**0,005**
wooden shoes	0	0	0		**0**	24	5	4,8	1,8	**0,013**	24	5	4,8	**0,003**
shoes for dancing on the table	33	5	6,6	3,5	**0,018**	17	2	8,5	0,7	**0,004**	50	7	7,1	**0,01**
slippers (shoes worn only inside the house)	27	5	5,4	2,3	**0,015**	30	6	5	2,6	**0,02**	57	11	5,2	**0,017**
shoes of the shape of a table (square or rectangular)	19	3	6,3	3,8	**0,006**	18	4	4,5	2,4	**0,008**	37	7	5,3	**0,007**
high-heel shoes	0	0	0		**0**	22	3	7,3	2,5	**0,007**	22	3	7,3	**0,002**
very big size of shoes	0	0	0		**0**	16	3	5,3	2,1	**0,005**	16	3	5,3	**0,001**
a shelf for shoes	19	3	6,3	1,2	**0,006**	8	3	2,7	2,1	**0,003**	27	6	4,5	**0,005**
shoes designed to be worn when walking on the table	10	3	3,3	0,6	**0,003**	3	1	3		**0.0003**	13	4	3,3	**0,001**
OPR	**0,367**					**0,536**					**0,518**			

	Science					Humanities					Total			
chair bread	SUM	NA	AI	SD	PR	SUM	NA	AI	SD	PR	SUM	NA	AI	PR
old bread (as tough as a chair)	0	0	0		**0**	36	6	6	2,8	**0,024**	36	6	6	**0,006**
bread that a person can eat only when he/she is sitting on a chair	10	2	5	0	**0,0022**	24	4	6	2,6	**0,0107**	34	6	5,7	**0,006**
bread which "sits" on the chair (bread put on the chair)	3	1	3		**0,0003**	15	4	3,8	3,1	**0,0067**	18	5	3,6	**0,003**
slow breakfast	19	2	9,5	0,7	**0,0042**	0	0	0		**0**	19	2	9,5	**0,001**
family reunion	18	2	9	1,4	**0,004**	0	0	0		**0**	18	2	9	**0,001**
bread in the shape of a chair	11	3	3,7	2,1	**0,0037**	19	3	6,3	3,1	**0,0063**	30	6	5	**0,005**
a chair in the shape of bread	12	2	6	0,6	**0,0027**	17	3	5,7	1,2	**0,0057**	29	5	5,8	**0,004**
dry lunch	0	0	0		**0**	14	3	4,7	2,5	**0,0047**	14	3	4,7	**0,001**
a type of bread	10	3	3,3	0,6	**0,0033**	9	3	3	2,6	**0,003**	19	6	3,2	**0,003**
OPR	**0,355**					**0,58**					**0,36**			

	Science					Humanities					Total			
car egg	SUM	NA	AI	SD	PR	SUM	NA	AI	SD	PR	SUM	NA	AI	PR
a car that looks like an egg	64	12	5,3	2,9	**0,0853**	75	12	6,3	2,3	**0,1**	139	24	5,8	**0,093**
an egg-shaped device for a car	0	0	0		**0.0006**	24	5	4,8	2,9	**0,0133**	24	5	4,8	**0,003**
a child's car-seat	9	1	9		**0,001**	25	3	8,3	0,6	**0,0083**	34	4	8,5	**0,004**
an egg delivered by a car	5	1	5		**0**	17	4	4,3	2,4	**0,0076**	22	5	4,4	**0,003**
a very small car	0	0	0		**0**	19	3	6,3	2,1	**0,0063**	19	3	6,3	**0,002**
a car that delivers eggs	18	4	4,5	1	**0,008**	13	2	6,5	2,1	**0,0029**	31	6	5,2	**0,005**
an egg with a hard shell (it does not break when carried by a car)	3	1	3		**0.0003**	7	3	2,3	1,5	**0,0023**	10	4	2,5	**0,001**
OPR	**0,905**					**0,82**					**0,912**			

	Science					Humanities					Total			
apple leg	SUM	NA	AI	SD	PR	SUM	NA	AI	SD	PR	SUM	NA	AI	PR
a disease causing the leg to look like an apple	13	2	6,5	3,5	**0,0029**	36	5	7,2	2,7	**0,02**	49	7	7	**0,01**
the stem of an apple	28	4	7	1,4	**0,0124**	30	4	7,5	1,7	**0,0133**	58	8	7,3	**0,013**
a fat leg	15	3	5	3,6	**0,005**	29	4	7,3	2,2	**0,0129**	44	7	6,3	**0,009**
a knee	0	0	0		**0**	25	3	8,3	0,6	**0,0083**	25	3	8,3	**0,002**
a healthy leg (with the quality of an apple - apples are healthy)	6	1	6		**0.0007**	18	3	6	4,4	**0,006**	24	4	6	**0,003**
an injured, sore leg	0	0	0		**0**	18	3	6	1,7	**0,006**	18	3	6	**0,002**
a robot constructed by *Apple*	0	0	0		**0**	16	3	5,3	1,2	**0,0053**	16	3	5,3	**0,001**
a type of a cake (in the shape of a leg, filled with apples)	11	2	5,5	0,7	**0,0024**	8	3	2,7	1,2	**0,0027**	19	5	3,8	**0,003**
OPR	**0,612**					**0,43**					**0,416**			

	Science					Humanities					Total			
street potato	SUM	NA	AI	SD	PR	SUM	NA	AI	SD	PR	SUM	NA	AI	PR
a person wandering in the streets	35	4	8,8	1	**0,0156**	87	12	7,3	2,3	**0,116**	122	16	7,6	**0,054**
food made of potatoes sold in the street	37	5	7,4	1,1	**0,0206**	51	7	7,3	2,8	**0,0397**	88	12	7,3	**0,029**
a homeless person	25	3	8,3	2,9	**0,0083**	62	9	6,9	2,8	**0,062**	87	12	7,3	**0,029**
a potato sold in the street	15	4	3,8	2,6	**0,0067**	54	8	6,8	2,8	**0,048**	69	12	5,8	**0,023**
a potato grown in a street	19	5	3,8	1,3	**0,0106**	32	6	5,3	3,3	**0,0213**	51	11	4,6	**0,016**
the worst quality potato	0	0	0		**0**	26	4	6,5	1,7	**0,0116**	26	4	6,5	**0,003**
a person who works on the street	0	0	0		**0**	20	4	5	2,9	**0,0089**	20	4	5	**0,002**
OPR	**0,44**					**0,513**					**0,482**			

	Science					Humanities					Total			
sky piano	SUM	NA	AI	SD	PR	SUM	NA	AI	SD	PR	SUM	NA	AI	PR
a beautiful piano song (heavenly beautiful)	43	6	7,2	3	**0,0287**	11	2	5,5	0,7	**0,0024**	54	8	6,8	**0,012**
a piano on the board of an airplane	12	3	4	3	**0,004**	10	1	10		**0,0011**	22	4	5,5	**0,002**
a blue piano (of the colour of sky)	12	2	6	4,2	**0,0027**	51	7	7,3	2,5	**0,0397**	63	9	7	**0,016**
a very good piano with great sound	12	2	6	2,8	**0,0027**	33	7	4,7	1,9	**0,0257**	45	9	5	**0,011**
a piano of a famous artist	3	1	3		**0.0003**	17	4	4,3	2,2	**0,0076**	20	5	4	**0,003**
a very big piano	0	0	0		**0**	22	3	7,3	2,5	**0,0073**	22	3	7,3	**0,002**
OPR	**0,811**					**0,54**					**0,404**			

	Science					Humanities					Total			
horse tree	SUM	NA	AI	SD	PR	SUM	NA	AI	SD	PR	SUM	NA	AI	PR
a big tree	30	4	7,5	2,9	**0,0133**	46	9	5,1	3,4	**0,046**	76	13	5,8	**0,027**
a tree in the shape of a horse	12	4	3	1,4	**0,0053**	59	9	6,6	2,9	**0,059**	71	13	5,5	**0,026**
a tree one can tie his/her horse on	6	1	6		**0.0007**	25	5	5	2,6	**0,0139**	31	6	5,2	**0,005**
a tree with fruits that horses like	0	0	0		**0**	19	4	4,8	2,4	**0,0084**	19	4	4,8	**0,002**
OPR	**0,714**					**0,496**					**0,471**			

	Science					Humanities					Total			
dog paper	SUM	NA	AI	SD	PR	SUM	NA	AI	SD	PR	SUM	NA	AI	PR
paper containing information about a dog	3	1	3		**0.0003**	42	5	8,4	0,9	**0,0233**	45	6	7,5	**0,008**
paper for dog's excrements	41	6	6,8	2,1	**0,0273**	34	5	6,8	1,3	**0,0189**	75	11	6,8	**0,023**
torn paper	0	0	0		**0**	35	5	7	2,9	**0,0194**	35	5	7	**0,005**
paper for recycling	8	2	4	1,4	**0,0018**	25	5	5	2,4	**0,0139**	33	7	4,7	**0,006**
paper with pictures of dogs	2	1			**0.0002**	26	3	8,7	1,5	**0,0087**	28	4	7	**0,003**
paper with low-quality contents written on it	0	0	0		**0**	17	3	5,7	2,5	**0,0057**	17	3	5,7	**0,001**
paper with 4 legs (in the shape of a dog)	3	1	3		**0.0003**	17	3	5,7	2,1	**0,0057**	20	4	5	**0,002**
toilet paper for dogs	29	4	7,3	4,3	**0,0129**	8	3	2,7	1,2	**0,0027**	37	7	5,3	**0,007**
wrinkled paper	15	3	5	3	**0,005**	9	2	4,5	3,5	**0,002**	24	5	4,8	**0,003**
OPR	**0,604**					**0,38**					**0,609**			

In *SKASE Journal of Theoretical Linguistics* [online]. 2012, vol. 9, no. 2 [cit. 2012-12-18]. Available on web page <http://www.skase.sk/Volumes/JTL21/pdf_doc/02.pdf>. ISSN 1339-782X.