

Analysis of Social and Expressive Factors of Requests by Methods of Text Mining

Daša Munková¹, Michal Munk¹, Zuzana Fráterová² and Beáta Ďuračková¹

¹Constantine the Philosopher University / Tr. A. Hlinku 1, 94974 Nitra, Slovakia

²University of Economics / Dolnozemská cesta 1, 85235 Bratislava, Slovakia

{dmunkova, mmunk}@ukf.sk, {zfraterova, bdurackova}@gmail.com

Abstract

In our paper we focus on analysing textual information usage (selected politeness factors of speech act) in mother tongue and in foreign language to identify phenomena of a language consciousness transfer from the mother tongue into a foreign language communication – transference phenomena – and their impact on textual structures of politeness in chosen languages. Our aim was to make an analysis of request texts written in English, Spanish and Slovak language, where we examined the occurrence of keywords, in our case the factors of politeness in mother tongue (Slovak) and in foreign languages (English and Spanish). We examined the formulation of requests made by two different groups, requests formulated by linguists - Slovak students studying English as their major subject - on one side, and the requests formulated by non-linguists - Slovak students studying Economy, with the knowledge of Spanish, - on the other side. We used cross-tabulation analysis and association rule analysis as our research methods. The findings are interesting mainly in terms of differences in the use of politeness factors in English and Slovak language, and also the concordance in the use of politeness factors in Slovak and Spanish texts of requests.

1 Introduction

One of the most important tasks of foreign language learning is to learn how to communicate with native speakers properly and fluently not only in routine but also in less common situations, so

that the foreign language communication sounds natural, that the students learn how to fulfil their communicative goals or are able to integrate into the life of a different culture.

This requires the development of awareness of the nature of language and its impact on the world (Svalberg, 2007).

Trompenaars (1998) called the culture as a common network of meanings. Different “cultural” meanings through the semantic codes are anchored in language and are created by the communication structures according to different principles and laws. One of these principles is politeness, which is examined by Pragmalinguistics. In pragmalinguistic language study, politeness communication represents one of the basic topics of successful implementation of language functionality and development of communicative competence (Hymes, 1996; Canale and Swain, 1980).

The politeness theory we used when examining production of speech acts of the requesters is the Brown and Levinson model (1987) that is, in various elaborated forms, still applicable today and forms the basis for newer models and definitions of politeness (Scollon and Scollon, 1995; Lim, 1994; Yabuuchi, 2006). Today, authors studying politeness rather focus on cultural relativity of politeness (Watts, Ide and Ehlich, 1992; Blum-Kulka, House and Kasper 1989; Wierzbicka, 1985) and on transition from examining static aspects of politeness to the dynamic ones. Older forms of static examining of politeness typically focused on speaker’s activity, speaker and listener’s image,

and on rules applied in production of politeness speech acts.

Learning to communicate in an additional language involves developing an awareness of the ways in which culture interrelates with language whenever it is used (Liddicoat, Papademetre, Scarino and Kohler, 2003; Hašková and Malá, 2008).

Each interlocutor creates his/her own unique speech acts (Cohen, 1996; Searle, 1979) and within them he/she uses the factors of politeness in various combinations and meanings. Since the level of foreign language acquisition is not on intermediate level (B1 or B2), the speaker (sender) usually simplifies his/her utterance in foreign language, applies utterances from his/her mother tongue or sometimes translates them (word by word) into foreign language, hence he/she cannot be aware of differences in meanings, which one and the same element can acquire in the other language.

We therefore believe that it is important to examine the rules of production of politeness speech acts, which the interlocutors use in the production of their spoken and written utterances in mother tongue as well as in foreign language.

Politeness communication involves various types of speech acts: a request, an apology, a complaint, an acknowledgement etc. A request is a communicative act whose aim is to achieve, through proper communicative tools, that the interlocutor fulfils a particular requirement. A request can take various forms depending on the relation between the interlocutors (if social power is present, a request can take the form of a command etc.). Usually, the interlocutor recognizes that the fulfilment of the request on his/her side is voluntary and its fulfilment is negotiated according to the way the request was formulated and what politeness factors were used. Consensually, the interlocutor, especially in situations when social power and social distance are present, tries to use common formulas and features (politeness patterns) to ensure “commonly” used requests, not to raise his/her partner’s doubts about his/her credibility by using a certain unusual communicative feature.

The graphic form of a politeness communication is a written text, mostly unstructured, providing various kinds of information exchanged between the sender and the receiver. It provides a large

amount of information, suitable mainly for a particular research or text mining. Text mining includes several research areas. Similarly to KDD (Knowledge Discovery in Databases) statistical methods and methods of machine learning are tools for data analysis in text mining (Hearst, 1999; Sullivan 2001). On the other hand, text mining builds mainly on theoretical and computational linguistics by data pre-processing (Neuendorf, 2002; Titscher et al, 2002; Hajičová, Panevová and Sgall, 2003; Weiss et al., 2005). The gist of text mining is processing of unstructured (textual) information, extraction of meaningful variables (turning words into numbers - meaningful indexes) from a text document, so that the information from the text can be used (made accessible) for various statistical methods and methods of machine learning. It allows us, for instance, to analyse the words or clusters of words used in a given text, their association or order, or to analyse whole texts in terms of determining similarities among them, relations among variables, or how the occurrence of one variable depends on others and so on. We can find some methods and applications in various research works (Maa, Sakagamia and Muratab, 2011), (Blache and Rauzy, 2011) and (Das and Bandyopadhyay, 2010; Stastny and Skorpil, 2007; Balogh, Magdin, Turcani and Burianova, 2011).

The order, association and variability of the factors of politeness are different in every language and culture, because they are based on different association rules in the given culture – based on a general but also an individual level.

The interlocutor has many features at his disposal to formulate a request, which are usually classified according to a specific structure (culturally given). According to Trosborg (1995), a request consists of internal and external features, thus its inner and basic part is its gist, a so called minimal unit, which can serve as the specific speech act. Its components are speaker’s or listener’s perspective, modality (a wish), direct vs. indirect request formulation, sentence and syntactical modifiers etc. Components that are added to the request gist (with different intensity according to the used features) and make its effect stronger are considered to be external features. Some of the external features are e.g. conversation opening sequences - greetings, appeals, attention getters (sorry, excuse me etc.); features that soften the request impact on decision making

(image/field) of the listener – external sequences such as commands, minimizers, explanations, asking for speaker’s agreement, pre-sequences, compliments, mitigating devices, politeness features, reducers, promises etc. (examples of which we introduce in the next chapter).

In our paper we focus on the analysis of the structure of collected unstructured texts of requests through a description of association rules found, which the Slovak students of English and Spanish language use in formulating requests. Within the structure of requests, we will try to find similarities and differences in the use of chosen social and expressive factors of politeness in the mother tongue (Slovak) and a foreign language (English and Spanish). The collected texts were formulated by intermediate students of English language, studying philological study programmes of English (teacher training or interpreting and translation studies), and intermediate students of Spanish language, non-philological study programme - Economy (level B1 or B2). This research sample was chosen to allow us to examine the transference phenomena in foreign language and to compare their characteristics (properties) in case of advanced and intermediate students in foreign language.

The rest of the paper is structured as follow. The next chapter deals with the methods of data pre-processing. We describe a particular corpora - text acquisition and information extraction from a text. The third chapter focuses on specific linguistic data analysis. At the end, we discuss the obtained results from the cross-tabulation analysis and association rules.

2 Methods

2.1 Corpora - Texts Acquisition

To obtain suitable information from text documents it is important, indeed essential to prepare and process data well. Tools like categorization, clustering and information extraction are used for data preparation (Feldman and Sanger, 2007). For instance, by proper categorization of documents, we can make the whole process of obtaining information easier (Paralič and Košťál, 2003).

If we want to do data mining, it is inevitable that the text has undergone several of following steps of pre-processing (Paralič et al., 2010):

1. Text conversion into an electronic form.
2. Cleaning of non-textual information, so-called conversion on plain text. To remove the non-textual information boards of tools in Java platform can be used.
3. Tokenization and segmentation. They belong to basic steps of text processing. Tokenization (Koehn, 2010) splits up the plain text into elementary units - tokens. By tokenization, we try to reduce text into sequence of tokens.
4. Lemmatisation and tagging. Porter’s algorithm is one of the most used algorithms for stemming of English words. For Slovak language this algorithm is not so much effective, since Slovak belongs to synthetic languages with a rich morphology. The best known algorithm, using the list of prefixes and suffixes which are separated from the token to obtain the stem in basic form, is minimal machine for stemming (Páleš, 1994), but it is complicated and computationally more demanding. For non-English languages, SnowballStemmer supporting also Spanish is very common. The next step of linguistic data pre-processing is tagging, which lies in assignment of grammatical tags.
5. Removing redundant, insignificant words, so-called stop words. These are words containing no significant information in texts.

By data pre-processing, it is important to take into account the following linguistic features:

Homonyms – words voiced or spelled in the same way, but having different meanings (*bat - animal; bat - baseball equipment; which/witch*). For the quality of text preparation, homonyms should be divided according to context into different terms, thereby their diversity will be ensured.

Synonyms – different words with the same or similar meaning (*beautiful, pretty, attractive*). For synonyms, it is advisable to integrate them under the same term, thereby the uniformity of meaning will be ensured.

Compounds – indicate one object, are made when two or more words are joined to form a new word. By separating, individual terms carrying different meaning are formed (*passport, grandmother, sister-in-law etc.*). For this reason, compounds should be included under one term.

In our case, we applied the above mentioned steps of data preparation from linguistic documents on texts of requests. These requests were obtained

from students studying Linguistics (linguists) and students studying Economy at university (non-linguists) with a B2 level of knowledge of foreign language, whether in electronic or handwritten form, as in their mother tongue (Slovak) so in a foreign language (English, Spanish). We classified the texts of requests into individual categories according to Díaz-Pérez (2003) and Trosborg (1995), who summarized the scenarios of speech acts.

2.2 Information Extraction from the Texts

Text sources in natural language offer lots of information, but not all of them are suitable for computational analysis. Though by using software for linguistic data preparation, large amounts of sources can be sorted out and useful information from the individual words, phrases or sentences can be extracted. Therefore the gist of information extraction is the identification of specific information, such as in our case, expressive and social factors.

Methods based on rules and statistical methods are used to identify specific information. The statistical methods are used by data of lower quality (e.g. information extraction from blogs etc.). The methods based on rules, which we also used in our case, are based on fixed characteristics under which they are generated (e.g. association or sequence rules). We chose them because they are appropriate for specific tasks such as extraction of social and expressive factors. If we want to extract them, we must have a defined list of social and expressive factors. In our case, we used classification of politeness factors in line with Trosborg (1995) and Díaz-Pérez (2003) and we defined the following 9 factors:

F1 Attention getter: combination of salutations, a form to express a social role: e.g. *addressing people (title, first name, last name, friendly appeal markers) and politeness factors*.

F2 Speaker's perspective: *could I, may I etc.*

F3 Listener's perspective: *can you, would you etc.*

F4 Politeness features: e.g. *thank you, please* immediately before or after the request core.

F5 Pre-sequences: elements before a request core.

F6 Post-sequences: features after the expressed request, usually it is explanation.

F7 Mitigating devices: features expressing an apology for disturbing.

F8 Minimizers: features minimising the impact of request.

F9 Compliments: features intensifying the likelihood of request fulfilment.

The first three represent social factors and the rest are expressive factors.

3 Linguistic Data Analysis

3.1 Cross-tabulation Analysis

In our case, a cross-tabulation analysis consists of two analyses. The first is an analysis of texts of requests formulated in mother tongue (Slovak) and in foreign language (English). These texts of requests were written by Slovak students studying Linguistics, whether within their teacher training or translation and interpreting studies. The second analysis includes texts of requests formulated in Slovak (mother tongue) and Spanish (foreign language). Texts of requests were obtained from students, non-linguists, who had learnt one foreign language during their own university studies and who had passed a basic language state exam (a level B2 of the Common European Framework of Reference for Languages).

With the help of the cross-tabulation analysis we investigated whether there is a difference in the use of various factors in Slovak and foreign language (English or Spanish).

	Chi-square	df	p
Pearson	114.9155	8	0.0000
Cont. coeff. C	0.2434		
Cramér's V	0.2509		

	Chi-square	df	p
Pearson	4.2681	8	0.8322
Cont. coeff. C	0.0412		
Cramér's V	0.0412		

Table 1. Results of cross-tabulation analysis a) Slovak vs. English b) Slovak vs. Spanish.

The only requirement (a validity assumption) of the use of chi-square test is a large amount of expected frequencies. The requirement is not violated, the expected frequencies $e_{ij} = r_i s_j / n$ are large enough (i.e. they are positive and not more than 20% of e_{ij} are less than 5, $e_{ij} > 34.36$). The contingency coefficient represents the degree of

dependency between two nominal variables. The value of coefficient (see Table 1a) is approximately 0.25, where 1 means perfect dependency and 0 means independency. There is a medium dependency between the occurrence of individual factors of politeness and the language in case of Slovak vs. English, the contingency coefficient is statistically significant. The zero hypotheses (see Table 1a) are rejected, which means that the occurrence (use) of individual factors of politeness depends on the language (Slovak or English).

In the second case (Slovak vs. Spanish), the contingency coefficient (see Table 1b) is approximately 0.04. Therefore, there is no dependency between the occurrence of individual factors of politeness and the language, the contingency coefficient is statistically insignificant. The zero hypotheses (see Table 1b) are not rejected, which means that the use of individual factors of politeness does not depend on language in case of Slovak vs. Spanish.

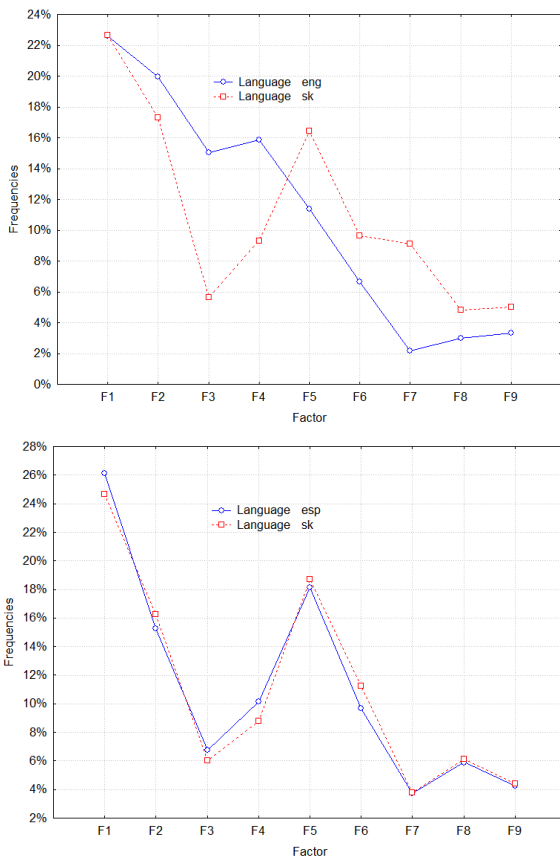


Figure 1. Interaction Plot - Language x Factor a) Slovak vs. English b) Slovak vs. Spanish.

The graph (Fig. 1a) shows the interaction frequencies *Language x Factor*. The graph presents a categorized polygon, where the factors of politeness are on the *x* axis and the observed frequencies of their usage (the occurrence) are on the *y* axis; while for each level of the variable *Language* one polygon is depicted. If the curves copy each other – they show the same course, the use of individual factors of politeness does not depend on the selected language. And vice versa, if there is any defined degree of dependency, the curves would not copy each other – which has been confirmed by the results of the analysis. We can observe different course for English and a different for Slovak. As we can see on the graph (Fig. 1a), the differences are mainly in factors F3, F4, F5 and F7. The factors F3 and F4 are considerably less used in Slovak than in English. Factor F3 (lis. perspective) represents a more direct and shorter utterance of a request. In terms of frequency, factor F2 (spe. Perspective) is much more preferred in the decision of perspective in mother tongue and also in foreign language. It means that an indirect utterance of a request and an attempt to avoid a direct addressing of requestee is more preferred. Factor F2 reduces the impact of a request, a requester, through the formulations (*May I borrow, copy ...*), takes over a part of “the effort” needed to fulfil the request upon him/herself, assuming, that the potential “alleviation” increases the likelihood of request fulfilment. The factor F4 is considerably less used in Slovak, that shows the requester’s knowledge of politeness structures in English requests with factor F4 (with words such as *please or thank you*) in comparison to Slovak. On the contrary, the factors F5 and F7 are much more often used in English. These are expressive factors. When the requester uses factor F5, he/she assumes that explaining the reasons to the requestee and requestee’s potential understanding of reasons of request may increase the likelihood of the fulfilment of a request. Consequently, the requester appeals to the empathy and imagination of the requestee, since he/she considers their influence as an effective strategy. Factor F7 (mitigating devices) reduces the impact of a request on the requestee, in terms of whether the requester does not interfere or over-interfere with his/her request in the requestee’s time, space or decision making.

The previous graph (Fig. 1b) visualises the interaction frequencies *Language x Factor* for Slovak and Spanish. In this case, the curves copy each other, they have the same course – the occurrence of individual factors of politeness does not depend on selected language, which is a confirmation of our analysis results. We can observe a similar course for Slovak as well for Spanish.

3.2 Association Rule Analysis

Similarly to cross-tabulation analysis, an association rule analysis is divided into two analyses - the analysis of requests written by linguists and the analysis of requests written by non-linguists.

The association rule analysis represents a non-sequential approach to the data being analysed. We will not analyze the sequences but transactions, so we will not include the order of factors used into the analysis. In our case, a transaction represents the set of factors observed in the texts of requests separately for English or Spanish and for Slovak.

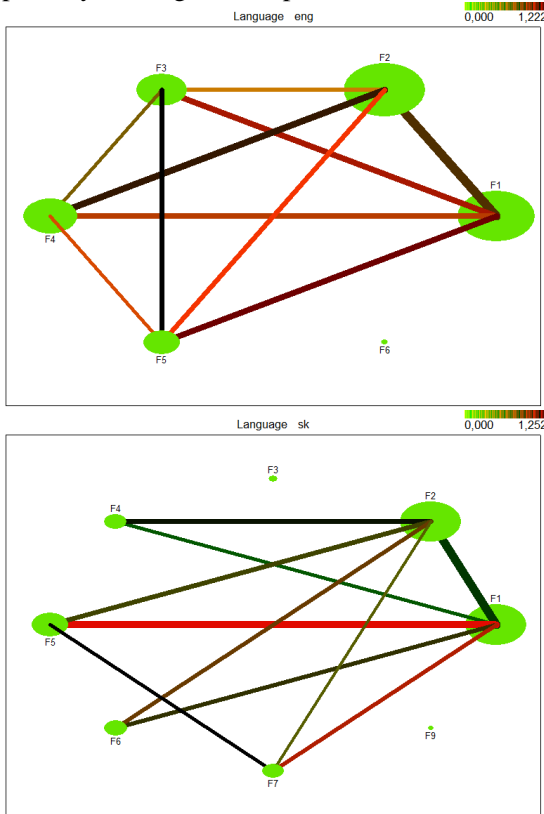
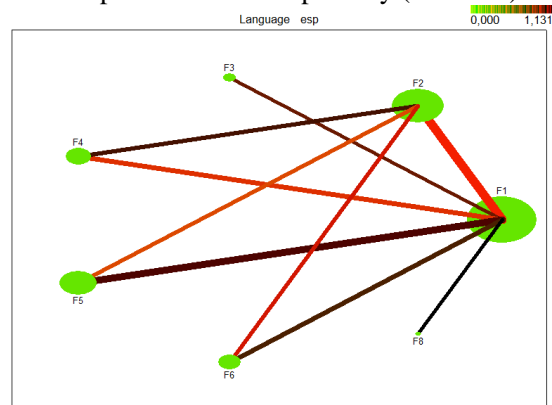


Figure 2. Web graph – a visualization of the discovered rules a) English b) Slovak.

The web graph (Fig.2a) depicts the discovered association rules for English requests, specifically the size of node represents the support of occurrence of the politeness factor, the thickness of the line represents the support of rule – pairs of factors (probability of occurrence in the pair) and the darkness of the line colour indicates a lift of the rule – the probability of a pair occurrence in transaction separately. We can see from the graph (Fig. 2a) that the factors of politeness F2, F1, F4 and F3 (support>51%) belong to the most frequently used factors. Similarly, like the combination of these factor pairs F1, F2; F2, F4, and F1, F3 (support>39%), the factors F5=>F3, F5=>F1, F2=>F4 and F1=>F3 occur in sets of factors of politeness more often together than as separate units (lift>1.11). In these cases the highest degree of interestingness was achieved – the lift, which defines how many times the selected factors of politeness occur more often together as if they were statistically independent. In case, that the lift is more than 1, the selected pairs occur more often jointly than separately in the set of used factors of politeness. It is necessary to take into account that in characterising the degree of interestingness – the lift, the orientation of the rule does not matter.

We found different association rules for Slovak requests than for English. The web graph (Fig. 2b) illustrates the discovered association rules. The most frequently used factors of politeness are F1, F2 and F5 (support>49%), as well as their pairs F1, F2 and F1, F5 (support>43%). The factors F7=>F5, F5=>F1, F4=>F2, F1=>F7 and F6=>F1 occur more often together in transactions of used factors of politeness than separately (lift>1.02).



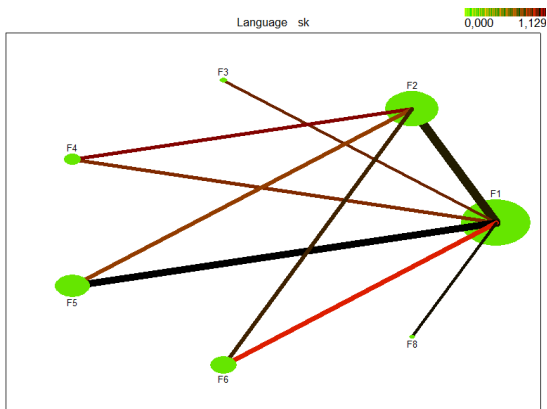


Figure 3. Web graph – a visualization of the discovered rules a) Spanish b) Slovak.

The web graph (Fig. 3a) visualizes the discovered association rules for requests written in Spanish. The graph (Fig. 3a) shows, that the factors of politeness F1, F2 and F5 (support>51%) belong to the most frequently used, similarly as the combinations of these couples of factors F1,F2 and F1,F5 (support>47%). The factors $F1 \Rightarrow F8$, $F5 \Rightarrow F1$, $F4 \Rightarrow F2$ and $F1 \Rightarrow F3$ occur in sets of used factors of politeness more often jointly than separately (lift>1.02).

We discovered almost identical association rules for texts of requests written in Slovak as those for Spanish. The previous graph (Fig. 3b) depicts the discovered association rules. The factors of politeness F1, F2 and F5 (support>51%) belong to the most frequently used, similarly as the combinations of these couples of factors F1, F2 and F1, F5 (support>48%). The factors $F5 \Rightarrow F1$, $F2 \Rightarrow F4$, $F1 \Rightarrow F8$ and $F2 \Rightarrow F1$ occur in transactions of used factors of politeness more frequently together than separately (lift>1.02).

The analysis results refer to the functioning of language consciousness of the requesters and the creation of politeness structure of utterance through the choice of factors. The politeness structure of Slovak has so far been investigated very peripherally. Therefore, in terms of comparison with Germanic and Romance languages this investigation is unique, and based on its results we can speculate not only about the decrease of transference regularities, but also about the politeness in Slovak language as such.

From our point of view, there are interesting pairs of expressive and social factors of politeness, i.e. mitigating device combined with pre-sequences but also with att. getter in a reverse order. It means

that, when a requester used an att. getter (a specific greeting etc.), it is more likely that he/she used an expressive factor, which raised the indirectness of the utterance and decreased its possible negative effect. Similarly, if he/she used indirect expression of perspective – F2 then he/she combined it with politeness features, so the most frequently observed association rules were those indicating the preference of indirect expression in Slovak.

4 Discussion and Conclusion

If we look at the results from the point of view of language used, in Slovak requests formulated by linguists the factors F1 (22.64%), F2 (17.30%) and F5 (16.46%) occurred most and the factors F8 (4.82%) and F9 (5.03%) the least frequently. In English requests, the factors F1 (22.62%), F2 (19.98%) and F4 (15.84%) occurred most frequently and factors F7 (2.18%), F8 (2.99%) and F9 (3.33%) least frequently.

The results of cross-tabulation analysis showed, that there is a difference between the language (Slovak or English) and the use of selected factors of politeness. This means that the occurrence of individual factors of politeness depends on the language used in the text of request.

We consider these findings interesting, because we examined the same requests but in different languages. Here, different patterns of request formulations are being created depending on the language used.

We presume that the level of English language acquisition influences in our case the use of politeness factors in requests and the concept, that the structure of politeness is different in target language than it is in mother tongue in case of factors F3, F4, F5 and F7. The requesters are aware of the differences, which weakens the possible transference of utterance and reduces the likelihood of errors in appropriateness of the utterance. Their utterance is simplified and more direct in the texts of requests in written English. We think, this is in order to ensure the understandability of their requests and is based on a well-known structure of politeness, which they know very well, so there is less risk of failure. In case of factors F1, F2, F8 and F9, they assume similar or the same usage in both languages and consciously do not think about (in)appropriateness of their frequency in foreign language, thus they

intensify the possible occurrence of errors caused by transference of consciousness of mother tongue into the foreign language.

The results of association rule analysis for texts of requests written in English showed, that the factors F2, F1, F4 and F3 (support: 71.24%; 68.58%; 53.98%; 51.77%) occurred most frequently among all factors of politeness in examined texts of requests.

The English requests are more direct with a politeness feature, which is a paradox. Linguists used much more often the lis. perspective (F3 for Slovak is 5.66% and for English 15.04%), and similarly also the politeness feature (F4 for Slovak is 9.33% and for English 15.84%), and considerably less pre-sequences (F5 for Slovak is 16.46% and for English 11.34%) and mitigating devices (F7 for Slovak is 9.12% and for English 2.18%), which are typical features of politeness in Slovak. The requester uses them to “ensure” the request fulfilment, which seems to be a successful strategy to approach the requestee and his/her understanding of the request. In English, their occurrence is less frequent.

In terms of factor combination, the following factors were combined the most: att. getter with spe. perspective, spe. perspective with politeness factor and att. getter with lis. perspective (support: 48.67%; 42.92%; 39.38%). From the point of view of pair occurrence $F5 \Rightarrow F3$, $F5 \Rightarrow F1$, $F2 \Rightarrow F4$ and $F1 \Rightarrow F3$ occurred more frequently jointly in transactions of used factors of politeness than as separate factors (lift: 1.22; 1.22; 1.12; 1.11).

In case of the couple pre-sequences \Rightarrow lis. perspective, the association of direct factors of politeness is shown. This means that when the requester used a pre-sequence, he/she also used the lis. perspective (to mitigate the directness of a request and its impact and effect on the listener). Pre-sequence and lis. perspective were associated with salutations and greetings (F5 with F1) or (F3 with F1) by requesters. They reinforce the request with them, i.e. they express the respect to the introductory - opening communication structures in the specific language and will not risk the failure of supposed communicated expectations of the partner – a native speaker. The next pair was spe. perspective and politeness feature (F2 with F4). In case when the author of English request used more direct utterance through factor F3, he/she mitigated this directness with expressive factor F4

(politeness feature). When he/she decided to express him/herself in a more indirect way, he/she used a combination with politeness feature (F2 with F4) reinforcing the likelihood of request fulfilment, which is confirmed by the last couple of factors.

The analysis results for Slovak requests were partially different. The most frequent factors used were: F1, F2 and F5 (support: 73.21%; 73.21%; 49.55%), contrary to English. As we mentioned before, Slovak language prefers indirect expressions with social factors of politeness that express the politeness model of requests in Slovak. Slovak expresses politeness through a more indirect utterance, explanation or compliments, and avoids interrupting the image of the communication partner, contrary to Spanish, which prefers a direct expression of request, considerably in the use of different expressive and language factors in request (as its politeness structure showed, the expressions of confidence – openness, directness are more preferred). The most frequent factor combinations are: att. getter with spe. perspective and att. getter with pre-sequences (support: 52.68%; 43.30%); and $F7 \Rightarrow F5$, $F5 \Rightarrow F1$, $F4 \Rightarrow F2$, $F1 \Rightarrow F7$ and $F6 \Rightarrow F1$ occur in transactions of used factors more frequently together than separately (lift: 1.25; 1.19; 1.16; 1.11; 1.02).

In Spanish requests written by students, whose major subject is not language, the following factors F1 (26.12%), F5 (18.14%) and F2 (15.27%) occurred most and factors F7 (3.72%), F9 (4.26%) and F8 (5.89%) least frequently. In requests formulated in Slovak, factors F1 (24.65%), F5 (18.69%) and F2 (16.24%) occurred most frequently and factors F7 (3.76%), F9 (4.41%) and F3 (6.04%) the least.

As we mentioned in chapter 3, no statistically significant difference between the used language (Slovak or Spanish) and chosen factors (contingency coefficient is 0.41) were proven. So it does not matter whether the requests are formulated in Slovak or Spanish, the requesters, students studying a non-philological subject, used the same factors of politeness.

Based on the differences in politeness structure of Spanish and Slovak language, we assumed that there would be differences in the use of factor F1 – considering other types of salutations and att. getter in both languages, differences in the use of

factors F2 and F3 - considering more direct expressions of requests in Spanish, and considerably lower factors F4, F7 and F9 used in Spanish language. Our assumption was not proven; in Spanish all factors of politeness have been fully applied in concordance with Slovak (general) but also with individual structure of politeness.

The results of the analysis showed for Spanish requests that factors F1, F2 and F5 (support: 82.00%; 65.00%; 51.33%) occurred among all the factors of politeness most frequently. In terms of factor pairs, att. getter with spe. perspective and att. getter with pre-sequences are used together most frequently (support: 53.00%; 47.00%). If we look at the factors in terms of couple occurrence, F1=>F8, F5=>F1, F4=>F2 and F1=>F3 occurred more often together in transactions of used factors of politeness than separately (lift: 1.13; 1.12; 1.06; 1.03; 1.02).

There is no point in discussing results for Slovak in detail because the results of association rule analysis were similar to those for Spanish language. Only one difference was shown in pair occurrence of post-sequences/explanation => spe. perspective and att. getter => lis. perspective occurred more frequently together than separately in Spanish and not in Slovak and vice versa, the couple lis. perspective => att. getter in Slovak and not in Spanish language.

We can say, that the requests in Slovak (the same in Spanish - considering the strong transference structure of these utterances) are less direct, using more mitigating devices (F7 - apologies for interference), such as I hope you don't mind me asking but could you read my outline and give some bibliographical references, please?; minimizers (F8), such as Please, can I borrow a book from the university library? I'll photocopy it and give it back to library next day, and compliments (F9) such as Excuse me, I know that you are a specialist on this and I asked myself if you could read my outline and if you could give me some bibliographical references., etc.

Partial differences between the use of factors of politeness of linguists and non-linguists are interesting. The linguists prefer factors F1, F2 and F5 in their mother tongue, combining them in varying degree and then complementing them with other expressive factors such as mitigating devices and post-sequences. Non-linguists add these three factors (although in a lower degree) to mitigating

devices, post-sequences and compliments. With their help non-linguists "ensure", to a higher degree, the request fulfilment by requestee.

The findings are interesting mainly in terms of differences in the use of politeness factors in English and Slovak, and also the concordance in the use of politeness factors in Slovak and Spanish requests formulations. Here we can see the impact of transference - a transfer of language awareness of native speaker in an utterance of foreign language mainly in case of students non-linguists, whose Spanish competency is at a lower level and they copy the usage of politeness factors without any knowledge and consideration of (in)appropriateness of their application in a given situation. The level of English competency of the linguists is higher and in case of factors F3, F4, F5 and F7, they choose different association rules, as well as the frequency of the use of individual factors. We assumed that some more complicated expressive factors (F5, F7, F8 and F9) would occur more frequently in foreign language, too. Students rather avoided them and they expressed themselves more directly in English or copied the Slovak politeness structure and "translated" their requests into another language without the awareness of its different politeness structure in Spanish. We assume that this could have been caused by uncertainty in foreign language use, mainly in English, but that is a focus of another research.

Acknowledgment

This work was supported by the Slovak Research and Development Agency under the contract No.APVV-0451-10.

References

- Agneta M-L. Svalberg. 2007. Language awareness and language learning. *Language Teaching*, 40:287-308.
- Alena Hašková and Eva Malá. 2008. Cudzí jazyk ako súčasť informačnej gramotnosti v zjednotenej Európe. In: *Informatika v škole 33/34*, 2008, s. 36 – 39. ISSN 1335-616X.
- Andrew D. Cohen. 1996. *Speech acts*. In *Sociolinguistics and Language Teaching*, Cambridge University Press, Cambridge.
- Anna Trosborg. 1995. *Interlanguage pragmatics: Requests, complaints, and apologies*. Mouton de Gruyter, Berlin.

- Anna Wierzbicka. 1985. Different cultures, different languages, different speech acts. *Journal of Pragmatics*, 9:145-178.
- Anthony J. Liddicoat et al. 2003. Report on intercultural language learning. Report to the Australian Government Department for Education Science and Training.
- A. Yabuuchi. 2006. Hierarchy politeness: What Brown and Levinson refused to see. *Intercultural Pragmatics*, 3(3):323-351.
- Dan Sullivan. 2001. Document Warehousing and Text Mining: Techniques for Improving Business Operations, Marketing and Sales. John Wiley & Sons.
- Dell H. Hymes. 1996. On communicative competence. In *The Communicative Approach to Language Teaching*, Oxford University Press.
- Dipankar Das and Sivaji Bandyopadhyay. 2010. Identifying Emotional Expressions, Intensities and Sentence level Emotion Tags using a Supervised Framework. In *PACLIC 24*, 95-104.
- Emil Páleš. 1994. SAPFO – Parafrázovač slovenčiny. VEDA, Bratislava.
- Eva Hajičová, Jarmila Panevová and Petr Sgall. 2003. Úvod do teoretické a počítačové lingvistiky. Karolinum, Praha.
- Fons Trompenaars. 1998. *Riding the Waves of Culture*. Nicolas Brealey, London.
- Francisco J. Diaz Pérez. 2003. La cortesía verbal en inglés y en español. *Actos de habla y pragmática intercultural*. Universidad de Jaén, Jaén.
- Ján Paralič and Ivan Košťál'. 2003. A Document Retrieval Method Based on Ontology Associations. *Journal of Information and Organizational Sciences*, 27:93-99.
- Ján Paralič et al. 2010. *Dolovanie znalostí z textov*. Equilibria, Košice.
- Jiri Stastny and Vladislav Skorpil. 2007. Genetic Algorithm and Neural Network. In *Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications*, 347-351.
- John R. Searle. 1979. *Expression and meaning: Studies in the theory of speech acts*. Cambridge University Press, Cambridge.
- Kimberly A. Neuendorf. 2002. *The Content Analysis Guidebook*. Sage, London.
- Marti A. Hearst. 1999. Untangling text data mining. In *ACL '99 Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, 3-10.
- Michael Canale and Merrill Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1:1-47.
- Penelope Brown and Stephen C. Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge University Press, New York.
- Philipp Koehn. 2010. *Statistical Machine Translation*. Cambridge University Press.
- Philippe Blache and Stéphane Rauzy. 2011. Predicting Linguistic Difficulty by Means of a Morpho-Syntactic Probabilistic Model. In *PACLIC 25*, 160-167.
- Richard J. Watts, Sachiko Ide and Konrad Ehlich. 1992. *Politeness in Language: Studies in its History, Theory, and Practice*. Mouton de Gruyter, Berlin/New York.
- Ron Scollon and Suzanne Wong Scollon. 1995. *Intercultural Communication: A Discourse Approach*. Blackwell, Oxford.
- Ronen Feldman and James Sanger. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- Sholom M. Weiss et al. 2005. *Text Mining: Predictive Methods for Analyzing Unstructured Information*. Springer.
- Shoshana Blum-Kulka, Juliane House and Gabriele Kasper. 1989. *Cross-cultural pragmatics: requests and apologies*. Ablex, Norwood.
- Stefan Titscher et al. 2002. *Methods of Text and Discourse Analysis*. Sage, London.
- Tae-Seop Lim. 1994. Facework and interpersonal relationships. In *The Challenge of Facework: Cross-Cultural and Interpersonal Issues*, 209-229.
- Qing Maa, Shinya Sakagamia and Masaki Muratab. 2011. Extraction of Broad-Scale, High-Precision Japanese-English Parallel Translation Expressions Using Lexical Information and Rules. In *PACLIC 25*, 577-586.
- Zoltan Balogh et al. 2011. Interactivity elements implementation analysis in e-learning courses of professional informatics subjects. In *8th International Conference on Efficiency and Responsibility in Education*, 5-14.