# How can 'rank distance' help in development of CALL material? A comparative study of needs analyses for producing e-material

**Anca Dinu[1], Liviu P. Dinu[2]**
**[1]University of Bucharest, Faculty of Foreign Languages, Edgar Quinet 5-7,**
**Bucharest, Romania, e-mail: anca_d_dinu@yahoo.com**
**[2] University of Bucharest, Faculty of Mathematics, Academiei 14,**
**Bucharest, Romania, e-mail: ldinu@funinf.cs.unibuc.ro**

Abstract:

The innovative aspect of this paper consists in using a mathematical method for describing the similarity between socio-cultural and linguistic environments of different host countries for exchange students. In this paper we process and compare the needs analyses, which have been produced for Czech, French, Polish, Portuguese and Romanian in the course of the development of the Euromobil2 project (a hybrid multimedia language training and information programme to support student mobility). The needs analyses reveal, among other things, problems that the exchange students are confronted with, such as unexpected features of the host countries (both in positive and negative sense) or differences in communicative strategies in study and everyday situations. We are interested in investigating the (dis)similarities between the results of the needs analyses for the five languages. We use the following strategy: For each language, the answers are categorized and the categories are then ranked according to their score. We compare the obtained rankings by using a particular method (called rank distance), which was successfully applied by the authors in different fields, such as computational linguistics, genomics and social sciences. Our research not only confirms and/or details some natural expectations, but offers some insights, otherwise inaccessible through direct observations or intuition. The results of this comparative study are expected to aid in the development of efficient e-material, specific for each target group.

**Keywords:** needs analyses, rank distance, Euromobil2, similarity, exchange students

## 1. Introduction

Student mobility today is supported by numerous EU funded programmes and bilateral contracts between educational institutions. The success of a study course abroad strongly depends on the language skills of the students and their knowledge of the culture and social conventions of their host country and host university. In order to offer exchange students the opportunity to prepare themselves efficiently for their stay abroad and to raise awareness for intercultural contacts, EUROMOBIL (www.euro-mobil.org), a Multimedia Language Learning and Information Programme on CD-ROM with links to

the web has been designed with the support of the European Commission (Socrates/ Lingua2, 225825-CP-1-2005-1-FR-LINGUA-L2).

Euromobil programmes consist of two main complementary components: tasks relating to different communicative situations in the target language and Information on target universities, countries and cultures.

The first part of the project, Euromobil1 had as target groups exchange student that were coming to study in Germany, England, Hungary and Finland.

More programmes to include French, Portuguese, Czech, Polish and Romanian are presently under development as part of the second part of the project, Euromobil2. The programme can be adapted for further languages.

Euromobil2 comprises four modules for each language: Studies, Library, Services and Leisure, each including a Glossary and a Good to know section.

## 2. Euromobil2 needs analyses

At the beginning of the Euromobil project, a needs analysis survey was conducted to plan the contents of the programme. Exchange students were asked to complete questionnaires in order to reveal their problems in each of the host countries. Thus, project partners could filter out the accidental problems from the current problems exchange students are confronted with.

The first version of the questionnaire was tested during Euromobil1 in Finland, and after that the questionnaire was changed to accommodate to specific features of the participating countries. Thus, slightly different questionnaires were used in Germany, Finland, and Hungary in the first year of the project. The results of these pilot version analyses served as a basis for developing the raw version of the EUROMOBIL. The concept of the German programme was applied when Great Britain joined the project during the second year.

In order to specify the needs and to develop a more accurate concept, the original questionnaire was radically shortened. It was given to the exchange student in Germany, Great Britain, Hungary and Finland during the second year of the project. The results of this survey were used particularly when creating the exercises for the language course, and in designing the contents of the Good to know section.

Euromobil2 questionnaire designing for needs analysis had benefited from all this experience gained during Euromobil1. The questionnaires were relatively short and a single common version was used for all five new countries. The project partners could thus compare the results of the needs analysis.

The idea was: given that most of the needs analyses answers are rankings, can we compare the problems exchange students were confronted with, while studying in a foreign country, and to draw a conclusion upon the similarities? And what formal method shall we use?

This article compares and interprets the data and results of Euromobil2 needs analysis using rank distance, a metric between rankings, which was already applied with good results in bioinformatics (the DNA similarity problem [Dinu and Sgaro, 2006]), computational linguistics (similarity of Romance languages [Dinu and Dinu, 2005]) and social sciences.

## 3. On the Similarities

In this section we analyze the data collected from the exchange students in each partner country. We selected some of the most informative questions from the questionnaires. We are interested in investigating the (dis)similarities between the results of needs analyses for the five languages, so that the project partners can develop e-material specific for each target group. For the complex data we use the following strategy: for each language, the answers are categorized and the categories are then ranked according to their score. We compare the obtained rankings by using a particular method called rank distance.

The mathematical tools are presented in mathematical addenda, at the end of the article: the rank distance, scaled total rank distance and a normalized distance to be used in cases in which two or more objects share the same position in a ranking.

The analyzed data are grouped into three categories: language, studying and socio-cultural similarities.

Our research not only confirms and/or details some natural expectations, but also offers some insights, otherwise inaccessible through direct observations or intuition.

## 3.1. Language problems similarities

The results of the student's answers at question A8 "Why did you want to become an exchange student? Do you want to:" are given in *table 1*:

| A8. Why did you want to become an exchange student? Do you want to: | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| learn the language of your host country | 62.50% | 88.00% | 50.77% | 51.00% | 59.46% |
| study a major subject | 39.58% | 60.00% | 40.00% | 34.69% | 35.13% |
| study a minor subject | 14.58% | 5.30% | 3.08% | 24.49% | 18.91% |

Table 1

As we can see in *table 1,* in all participating countries the motivations of the students were similar as classification, the most important being to learn the language of the host country, then to study a major subject and then to study a minor subject, in this order. This shows a particularly strong interest of all students in learning the language of the host country. As a consequence, all project partners included in the application consistent parts for learning the host country language, with interactive tasks, vocabulary and grammar.

The results of the student's answers at question B1 "Please evaluate your language proficiency on your arrival in the host country" are given in *table 2*:

| B1. Please evaluate your language proficiency on your arrival in the host country | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| None | 68.75% | 5.30% | 30.77% | 40.82% | 40.54% |
| (very) poor | 18.75% | 21.30% | 33.85% | 42.86% | 24.32% |
| fair/satisfactory | 8.33% | 53.30% | 15.38% | 12.24% | 16.22% |
| Good | 0.00% | 17.30% | 16.92% | 2.04% | 8.11% |
| Fluent | 2.08% | 0.00% | 3.08% | 2.04% | 5.41% |

Table 2

In order to interpret these results, we extracted from *table 2* a ranking of levels of proficiency arranged in descending order of percentage of students' answers, for each country. For the sake of simplicity we denoted None by N, (very) Poor by P, Fair/Satisfactory by S, Good by G and Fluent by F. *Table 3* shows these rankings:

| CZ | FR | PL | PO | RO |
|---|---|---|---|---|
| N | S | P | P | N |
| P | P | N | N | P |
| S | G | G | S | S |
| F | N | S | G and F | G |
| G | F | F | | F |

Table 3

We want to find out how similar are these rankings w.r.t. each other. Thus, in order to interpret these data, we need a tool for comparing rankings. We choose to use scaled total rank distance (STRD).

We summarize the computed distances in the *table 4* (which is a symmetric matrix):

| normalized STRD | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| **CZ** | 0.00 | 0.60 | 0.43 | 0.36 | 0.03 |
| **FR** | 0.60 | 0.00 | 0.53 | 0.43 | 0.53 |
| **PL** | 0.43 | 0.53 | 0.00 | 0.09 | 0.37 |
| **PO** | 0.36 | 0.43 | 0.09 | 0.00 | 0.36 |
| **RO** | 0.03 | 0.53 | 0.37 | 0.36 | 0.00 |

Table 4

We can see that every time FR finds itself at the biggest distance from every other language. This is due to the fact that the exchange students coming to France are more advanced in studying the language of the host country than all the others. This supports the decision of the project partners to design for France material for advanced level and for the other countries for beginner's level.

The results of the student's answers at question B2 "To what extent do you have difficulties in the following areas?" are given in *table 5*:

| B2. To what extent do you have difficulties in the following areas? | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| understanding the speech of teachers and room mates | 76 | 139 | 96 | 78 | 62 |
| understanding complex sentences | 90 | 126 | 100 | 95 | 72 |
| understanding the content | 77 | 111 | 102 | 65 | 58 |
| understanding the connecting thought | 75 | 120 | 113 | 66 | 65 |
| participating in a discussion | 97 | 166 | 111 | 105 | 74 |

Table 5

The students were asked to give scores from 0 to 4 (0 = none, 1 = a little, 2 = moderate, 3 = a lot, 4 = very much) to each of the five categories problems. For each country, we computed the general score for each category of problems as the sum of number of answers ponderated with the score.

*Example*: For PO, the general score for the category of problem "*understanding the speech of teachers and room mates*" was computed as $0*7+1*16+2*20+3*6+4*1 = 78$.

| | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| understanding the speech of teachers and room mates | 7 | 16 | 20 | 6 | 1 |

Thus, the bigger the general score for a category of problems, the more significant are the difficulties of the exchange students confronted with this category of problem.

To analyze these results, we proceed in a similar manner as for the analysis of the previous B1 question. From *table 5* we obtain for each country a ranking of category of problems arranged in descending order of the general score (*table 6*). We denoted *understanding the speech of teachers and room mates* by S, *understanding complex sentences* by CS, *understanding the content* by C, *understanding the connecting thought* by CT and *participating in a discussion* by PD.

| CZ | FR | PL | PO | RO |
|---|---|---|---|---|
| PD | PD | CT | PD | PD |
| CS | ST | PD | CS | CS |
| C | CS | C | ST | CT |
| ST | CT | CS | CT | ST |
| CT | C | ST | C | C |

Table 6

We summarize the computed distances in table 7:

| normalized STRD | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| **CZ** | 0.00 | 0.24 | 0.38 | 0.10 | 0.10 |
| **FR** | 0.24 | 0.00 | 0.57 | 0.13 | 0.20 |
| **PL** | 0.38 | 0.57 | 0.00 | 0.63 | 0.57 |
| **PO** | 0.10 | 0.13 | 0.63 | 0.00 | 0.07 |
| **RO** | 0.10 | 0.20 | 0.57 | 0.07 | 0.00 |

Table 7

We can see that except for PL, for the other four languages (CZ, FR, PO and RO) the differences between their rankings are very small, between 0.07 and 0.24. We remember that the distance is normalized (all possible values range from 0.00 to 1.00). This shows that exchange students have very similar difficulties w.r.t. the language problems mentioned in question B2.

The results of the student's answers at the question B3 "In which situations have you had communicative problems?" are given in *table 8*:

| **B3. In which situations have you had communicative problems?** | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| finding/renting a flat | 51 | 55 | 82 | 65 | 56 |
| at the cafeteria/in the restaurant | 58 | 47 | 74 | 46 | 36 |
| at the doctor/dentist | 29 | 60 | 102 | 34 | 31 |
| at the hairdresser | 41 | 59 | 105 | 30 | 36 |
| in shops | 69 | 61 | 89 | 58 | 35 |
| at the post office | 66 | 43 | 110 | 43 | 48 |
| at the bank | 43 | 98 | 82 | 50 | 52 |
| at the travel agency | 24 | 41 | 59 | 19 | 33 |
| at the railway/bus station | 66 | 46 | 103 | 45 | 30 |
| at the museum | 33 | 29 | 77 | 32 | 0 |
| at the theatre/movie/in concert | 35 | 51 | 69 | 42 | 52 |
| at the swimming pool | 35 | 26 | 56 | 27 | 31 |
| when renting equipment (bike, ski…) | 25 | 34 | 80 | 30 | 31 |
| in following media (radio, TV, newspapers) | 120 | 129 | 137 | 55 | 62 |

Table 8

Again, the students were asked to give scores from 0 to 4 (0 = none, 1 = a little, 2 = moderate, 3 = a lot, 4 = very much) to each of categories of communicative situations. We computed the general score for each category as for B2 question.

We denoted *finding/renting a flat* by F, *at the cafeteria/in the restaurant* by C/R, *at the doctor/dentist* by D, *at the hairdresser* by H, *in shops* by S, *at the post office* by PO, *at the bank* by B, *at the travel agency* by TA, *at the railway/bus station* by R, *at the*

*museum* by M, *at the theatre/movie/in concert* by T/M/C, *at the swimming pool* by SP, *when renting equipment* by RE and *in following media* by FM.

The rankings of the communicative situations arranged in descending order of the general score are given in table 9:

| CZ | FR | PL | PO | RO |
|-----|-----|-------|-------|-------|
| FM | FM | FM | F | FM |
| S | B | PO | S | F |
| PO | S | H | FM | B |
| R | D | R | B | T/M/C |
| C/R | H | D | C/R | PO |
| F | F | S | R | C/R |
| B | T/M/C | F | PO | H |
| H | C/R | B | T/M/C | S |
| T/M/C | R | RE | D | TA |
| SP | PO | M | M | D |
| M | TA | C/R | H | SP |
| D | RE | T/M/C | RE | RE |
| RE | M | TA | SP | R |
| TA | SP | SP | TA | M |

Table 9

Because more than one category occupies the same place in some of the rankings in *table 9*, we used a normalized distance related to rank distance, which can cope with this situation. We will keep using this distance from now on when confronted with such situations.

We summarize the computed distances in *table 10*:

| normalized distance | CZ | FR | PL | PO | RO |
|-----|------|------|------|------|------|
| **CZ** | 0.00 | 0.35 | 0.29 | 0.20 | 0.27 |
| **FR** | 0.35 | 0.00 | 0.34 | 0.29 | 0.35 |
| **PL** | 0.29 | 0.34 | 0.00 | 0.36 | 0.47 |
| **PO** | 0.20 | 0.29 | 0.36 | 0.00 | 0.32 |
| **RO** | 0.27 | 0.35 | 0.47 | 0.32 | 0.00 |

Table 10

As we can see from these results, the distances between the rankings of the communicative situations in the five countries are very similar to one another and faire small, between 0.29 and 0.47. This means that the communicative situation in which the exchange students have difficulties are pretty similar for all participating countries and the developers should collaborate during the development of the material, but also should stress the specific communicative problems for each country according to the results of the needs analysis.

## 3.2. Studying forms similarities

From the answers to question C3 one can see that most of the exchange students have founded some different forms of studying in the host university (*table 11*).

| C3. Were you used to different forms of studying at your home university | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| Yes – some are different here | 54.17% | 74.70% | 41.54% | 48.98% | 36.73% |
| No – all same as at home | 37.50% | 14.70% | 44.62% | 26.53% | 20.41% |
| Yes – all are different here | 6.25% | 9.30% | 9.23% | 22.45% | 21.62% |

Table 11

In order to better prepare the exchange students for the usual forms of study and examination of the host country, question C2 determines exactly which forms of academic activities exchange students founded in each host country (*table 12 and 13*).

| C2. What forms of academic activity are you involved in at your host university? | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| Lectures | 81.25% | 78.67% | 81.54% | 28.57% | 48.65% |
| Seminars | 54.17% | 32.00% | 61.54% | 32.65% | 24.32% |
| Practical training | 52.08% | 61.33% | 46.15% | 65.31% | 5.41% |
| Laboratory work | 22.92% | 61.33% | 40.00% | 32.65% | 5.41% |
| Written examinations | 64.58% | 84.00% | 61.54% | 75.51% | 21.62% |
| Oral examinations | 54.17% | 53.33% | 49.23% | 32.65% | 24.32% |
| Language courses | 81.25% | 70.67% | 52.31% | 55.10% | 27.03% |

Table 12

The rankings of the forms of study and examination arranged in descending order of the general score are given in *table 13*:

| CZ | FR | PL | PO | RO |
|---|---|---|---|---|
| L | W | L | W | L |
| LC | L | S | PT | LC |
| W | LC | W | LC | S |
| S | PT | LC | S | O |
| O | LW | O | LW | W |
| PT | O | PT | O | PT |
| LW | S | LW | L | LW |

Table 13

We give the computed distances in *table 14*:

| normalized distance | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| **CZ** | 0.00 | 0.36 | 0.21 | 0.61 | 0.17 |
| **FR** | 0.36 | 0.00 | 0.45 | 0.40 | 0.54 |
| **PL** | 0.21 | 0.45 | 0.00 | 0.67 | 0.25 |
| **PO** | 0.61 | 0.40 | 0.67 | 0.00 | 0.79 |
| **RO** | 0.17 | 0.54 | 0.25 | 0.79 | 0.00 |

Table 14

It is hard to notice a pattern in table 13 just by looking at it.

But after we have computed the values of the distances between the forms of academic activities from each country, we can distinguish two clusters of countries w.r.t. the similarities of the forms of academic activities: CZ, PL and RO on the one hand (with values of the distance between them of just 0.17, 0.21 and 0.25) and FR and PO on the other (the distance from east European countries to FR and PO varies from 0.54 to 0.79, the maximum possible value being again 1). One possible explanation for these results could be the regional pattern of teaching: more lectures for the first group and more student activities for the second.

This means that the project developers should include in the e-material information about the system of teaching and examination *specific* for each group of countries (or even for each country).

## 3.3. Socio-cultural and everyday situations similarities

The percentage of students that broke some cultural convention is relatively small for all countries. It is remarkable thaw that one in four exchange students studying in France founded themselves in a funny or embarrassing situation because they broke a cultural convention (*table 15*).

| **D7. Have you ever found yourself in a situation in the host country that was funny or embarrassing because you broke a cultural convention?** | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| yes | 16.67% | 25.30% | 10.77% | 2.04% | 8.11% |
| no | 77.08% | 64.00% | 87.69% | 97.96% | 83.78% |

Table 15

As for difficulties when dealing with practical matters, as administrative or everyday situations, the categories of problems are ranked quite different (*table16 and 17*).

| B6. Have you had any difficulties in dealing with practical matters or getting to know the university as? | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| Registration | 27.08% | 42,7% | 3.08% | 21.24% | 40.54% |
| Housing | 8.33% | 20,0% | 4.62% | 10.20% | 2.70% |
| Banking services | 4.17% | 18,7% | 3.08% | 8.16% | 13.51% |
| Finding your way around the campus | 14.58% | 16,0% | 4.62% | 2.04% | 0.00% |
| Using the library | 20.83% | 12,0% | 9.23% | 2.04% | 10.81% |
| Computing center services | 25.00% | 17,3% | 1.54% | 10.20% | 13.51% |
| Health care services | 4.17% | 6,7% | 4.62% | 21.24% | 0.00% |
| Meals | 16.67% | 8,0% | 7.69% | 2.04% | 8.11% |

Table 16

| CZ | FR | PL | PO | RO |
|---|---|---|---|---|
| R | M | L | R | R |
| C | HC | M | HC | B |
| L | R | H | H | C |
| M | H | F | C | L |
| F | C | HC | B | M |
| H | B | R | F | H |
| B | L | B | L | F |
| HC | F | C | M | HC |

Table 17

We give the computed distances in *table 18*:

| Distance normalized | CZ | FR | PL | PO | RO |
|---|---|---|---|---|---|
| CZ | 0.00 | 0.50 | 0.41 | 0.53 | 0.25 |
| FR | 0.50 | 0.00 | 0.53 | 0.35 | 0.47 |
| PL | 0.41 | 0.53 | 0.00 | 0.78 | 0.79 |
| PO | 0.53 | 0.35 | 0.78 | 0.00 | 0.48 |
| RO | 0.25 | 0.47 | 0.79 | 0.48 | 0.00 |

Table 18

We can only remark that the lowest distance between the difficulties when dealing with practical matters is reached for CZ and RO. Practically, except banking services, which are more problematic for exchange students in RO then in CZ, the rankings for the two countries are quasi the same.

For the rest of the countries, the distances are quite high, so the developers of e-materials have to include information about dealing with practical matters or getting to know the university, independent of each other, according to the specific priorities for each country, as they emerge from needs analysis.

## 4. Conclusions

EUROMOBIL, a Multimedia Language Learning and Information Programme on CD-ROM with links to the web has been designed with the support of the European Commission (Socrates/ Lingua2), in order to offer exchange students the opportunity to prepare themselves efficiently for their stay abroad and to raise awareness for intercultural contacts,.

At the beginning of the Euromobil project, a needs analysis survey was conducted to plan the contents of the programme. Exchange students were asked to complete questionnaires in order to reveal their problems in each of the host countries. Thus, project partners could filter out the accidental problems from the current problems exchange students are confronted with. The analyzed data were grouped into three categories: language, studying and socio-cultural similarities.

The idea was: given that most of the needs analyses answers are rankings, can we compare the problems exchange students were confronted with, while studying in a foreign country, and to draw a conclusion upon the similarities? And what formal method shall we use?

This article compares and interprets the data and results of Euromobil2 needs analysis using rank distance, a metric between rankings.

Our research not only confirms and/or details some natural intuitions, but offers some insights, otherwise inaccessible through direct observations.

The results of this comparative study are expected to aid in the development of efficient e-material, specific for each target group, during Euromobil2 project.

## 5. Mathematical addenda

A ranking is an ordered list of objects
Let $u=x_1x_2\ldots x_n$ and $v=y_1y_2\ldots y_m$ be two rankings of lengths n and m, respectively.
For an element $x_i$ in u we define its order or rank by **$ord(x_i \mid u)=n+1-i$**,
i.e., the rank of $x_i$ is its position in ranking, counted from the **bottom to the top**.
Rank Distance [Dinu 2003, Dinu 2005] between two rankings u and v is given by:

$$\Delta(u,v) \quad = \sum_{x \in u \cap v} |ord(x|u) - ord(x|v)| +$$
$$+ \sum_{x \in u \setminus v} ord(x|u) + \sum_{x \in v \setminus u} ord(x|v)$$

The main idea of rank distance is that it penalizes the unmatched objects (i.e. founded just in one of the rankings) more if they are situated in the initial part of the rankings.

Scaled total rank distance [Dinu and Dinu, 2006] was created to also penalize the matched objects more if they are situated in the initial part of the rankings.

Scaled total rank distance between two rankings *A* and *B* of length n is:

$$S(A, B) = \sum_{i=1}^{n} \frac{\Delta(A_i, B_i)}{i(i+1)}.$$

where $A_i$ and $B_i$ are partial rankings formed by keeping the first *i* elements from the rankings *A* and *B*, respectively.

In situations where more than one category occupies the same place in a ranking, we used a normalized distance related to rank distance, which can cope with such situations; this distance is given by the formula:

$$\mathbf{D}(u, v) = \sum_{x \in u \cap v} |ord(x|u) - ord(x|v)| \times \mathbf{max}\left(\mathbf{ord(x|u), ord(x|v)}\right) + $$
$$ + \sum_{x \in u \setminus v} ord(x|u) + \sum_{x \in v \setminus u} ord(x|v)$$

**REFERENCES**

[1] L.P. Dinu, 2003 *On the classification and aggregation of hierarchies with different constitutive elements*, Fundamenta Informaticae, 55(1), 39-50, 2003.

[2] L.P. Dinu, *Rank distance with applications in similarity of natural languages*, Fundamenta Informaticae, 65(1-4), 135-149, 2005.

[3] A. Dinu, L.P. Dinu, 2005. *On the Syllabic Similarities of Romance Languages*. Lecture Notes in Computer Science, Volume 3406, pp. 785-789, 2005

[4] A. Dinu, L. P. Dinu, 2006. *Total rank distance and scaled total rank distance: two alternative metrics in computational linguistics.* In J. Nerbonne & E.Hinrichs (eds.) Proceedings *Linguistic Distances* Workshop at the joint conference of International Committee on Computational Linguistics and the Association for Computational Linguistics (ACL-COLING 2006), Sydney, July, 2006, pp109-117.

[5] L P. Dinu, A. Sgarro, 2006. *A low-complexity distance for DNA strings* Fundamenta Informaticae 73(3), 361-372, 2006