# Choosing the Best Comparison Under the Circumstances

Amy Isard

Language Technology Group, School of Informatics, University of Edinburgh
amy.isard@ed.ac.uk

**Abstract.** The Methodius Natural Language Generation systems generates personalized descriptions of objects from a collection. As part of the user modeling component, it creates comparisons between the current object being viewed and previous objects from the user history. We present our general algorithm for choosing the best comparison, which can be optimized to give the best result for different domains through a parameterizable scoring function.

## 1 Introduction

The Methodius Natural Language Generation (NLG) system creates personalized descriptions of objects from a database which can be displayed in a variety of modalities, including a virtual museum on the web, or in a real museum setting, through text or speech on a handheld device or through dialogue with a robot museum guide. While a user navigates through a domain, comparisons can be made between an object they are currently viewing and those which have come before. This paper describes a novel algorithm for selecting the most relevant and interesting comparisons in the context.

We first give an overview of Natural Language Generation (section 2), and then motivate the use of Comparisons in Cultural Heritage description systems (section 3). We then describe the new Comparison Algorithm in detail (section 4) and present an example of the algorithm in action (sections 5 and 6) and finally some conclusions and future directions (section 7).

## 2 Natural Language Generation

The Methodius system is a descendant of the Exprimo generation system developed during the M-PIRO project [1], which generated texts about ancient Greek artefacts selected by curators at the Foundation of the Hellenic World[1]. A web interface allowed users to navigate through the collection by clicking on thumbnail images of the objects. Methodius is designed to be a more robust and modular system, which can deal with collections of at least a million objects, and can be used for any domain in which an ontology of objects and attributes exist. We currently have a test domain of built heritage sites from the Royal Commission for the Ancient and Historical Monuments of Scotland (RCAHMS)[2], and another from an online radio station, with descriptions of songs.

[1] http://www.fhw.gr

[2] http://www.rcahms.gov.uk

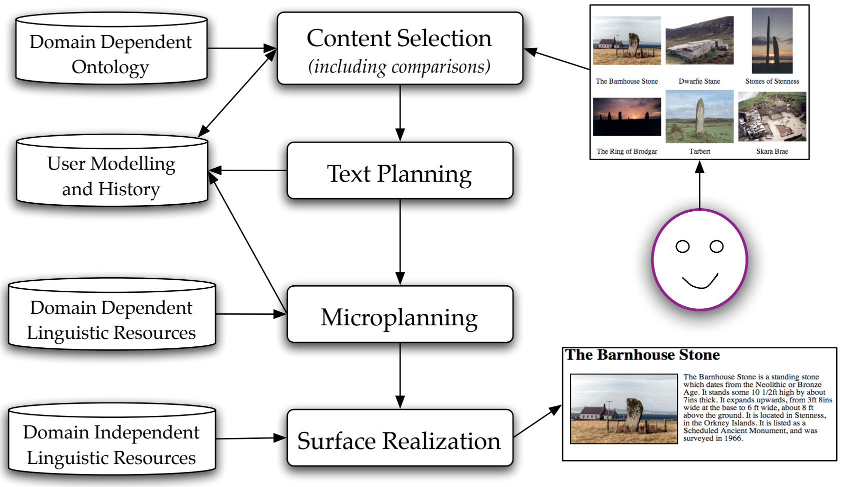

Domain Dependent Ontology
Content Selection
(including comparisons)
User Modelling and History
Text Planning
Domain Dependent Linguistic Resources
Microplanning
Domain Independent Linguistic Resources
Surface Realization
The Barnhouse Stone
Dwarfie Stane
Stones of Stenness
The Ring of Brodgar
Tarbert
Skara Brae
The Barnhouse Stone
The Barnhouse Stone is a standing stone which dates from the Neolithic or Bronze Age. It stands some 10 1/2ft high by about 7ins thick. It expands upwards, from 3ft 8ins wide at the base to 6 ft wide, about 8 ft above the ground. It is located in Stenness, in the Orkney Islands. It is listed as a Scheduled Ancient Monument, and was surveyed in 1966.

**Fig. 1.** Output Text: The Barnhouse Stone is a standing stone which dates from the Neolithic or Bronze Age. It stands some 10 1/2ft high by about 7ins thick. It expands upwards, from 3ft 8ins wide at the base to 6 ft wide, about 8 ft above the ground. It is located in Stenness, in the Orkney Islands. It is listed as a Scheduled Ancient Monument, and was surveyed in 1966.

The system uses a typical NLG architecture based on the pipeline model described in [2]. Figure 1 shows how the generation component could fit into a whole system. In this example, based on the RCAHMS data, a user clicks on a thumbnail picture of the Barnhouse Stone, a standing stone from the island of Orkney in Scotland, and is taken to a web page with a larger picture of the stone and a paragraph of text generated from the underlying database. The full descriptive text for the Barnhouse Stone is shown in the caption of figure 1.

The first phase of the generation is Content Selection where an algorithm is used to select a subset of the available facts about the object which has been chosen by the user, based on user modelling information. Rhetorical structure is then used to group and reorder the sentences during the Text Planning, before Microplanning and Surface Realization instantiate the texts using pre-defined linguistic templates.

As the user navigates through the thumbnail pictures, it becomes possible for the system to make comparisons between the object which has just been selected, and those which have already been viewed. This paper focusses on the comparison algorithm, which is part of the content selection phase, when the system is deciding which information abuot the current object to present to the user.

## 3 Why Add Comparisons?

A number of previous natural language generation systems have included comparisons, which can improve the clarity of texts which describe objects in a hierarchy, and can give users a better view of a domain by making differences and similarities explicit. Comparisons can also facilitate learning by relating new concepts to a user's existing knowledge, and by repeating pieces of information, to allow reinforcement learning.

### 3.1 Comparisons in Different Domains

Foster [3] provides a useful review of previous systems which use comparisons, but only those which produced descriptions in cultural heritage domains will be mentioned here. ILEX [4] generated descriptions of items of jewellery in a collection at the National Museum of Scotland, and POWER [5] used data from the Powerhouse Museum in Sydney. Both of these used comparison modules adapted from PEBA-II [6, 7], a system which generated hypertext descriptions of animals in a Linnean taxonomy. In all of these systems, as user navigates through a hypertext representation of the domain. In POWER and PEBA-II, the user could request a comparison between two objects (a direct comparison) and all three systems generated comparisons between the object currently being viewed, and a previous object from the user history (illustrative and clarificatory comparison).

### 3.2 Referring to Objects and Groups of Objects

The systems described in section 3.1 use various algorithms to decide which previous object(s), and which attributes of the object(s), should be selected for comparison with the object currently in focus. An important factor in this decision is the possibilities available for referring to previous objects. In some domains, the objects have names, whereas in others, they must be referred to by their type, possibly disambiguating with attribute references. For example, in POWER, the objects can be referred to by name (e.g. The Analytical Engine), whereas in ILEX and M-PIRO they are described by type (e.g. the necklace, the amphora). The comparison algorithms deal with this issue in different ways. In ILEX, objects are distinguished by particular attributes, for example "Like the necklace designed by Flockinger, this item is in the organic style."

In M-PIRO, objects are not disambiguated in this way; comparisons are made either with the previous object, or with a group of previous objects which share the same type. Unlike the previous systems, which only allowed comparisons between single objects, M-PIRO also generated comparisons between the current object being viewed and groups of objects from the user history [8]. Therefore, in addition to the direct, illustrative and clarificatory comparisons, this also allowed meaningful contrastive comparisons such as "This exhibit is another stater. Unlike the previous staters, which are made of silver, this stater is made of gold."

The algorithm being proposed here supports both referring by name, and referring by type, thus allowing for flexible comparisons in domains in which either or both sorts of object are represented.

## 4 Methodius Comparison Scoring Heuristics

The comparison module of the Methodius system makes comparisons between the currently selected object, known as the focal object (FO), and one or more previously-viewed objects, known as the comparison group (CG).

We use the following heuristics (H) and features (F) in choosing the best comparison:

**H1** Comparisons are more meaningful if made with a group with a larger number of members
**F1** The number of members in a CG

**H2** Comparisons are more meaningful if more can be made with the same comparator(s)
**F2** The number of comparisons which can be made between the FO and a CG

**H3** Comparisons with more closely related entities are more interesting
**F3** The hierarchical distance between the FO and the CG[3]

**H4** Comparisons with more recently viewed entities are more salient
**F4** The historical distance between the FO and the most recent member of the CG

**H5** A user will gradually forget details of previous objects
**F5** A limit to the number of previous objects which will be considered for comparison

In different domains and display situations, each of these features may be more or less important. We therefore propose an algorithm in which the features are each given a weight. This parameterization allows flexibility in order to allow experimentation to establish the best weights to be given to each in a particular domain. The comparison score equation is shown in equation 1.

$$(\alpha \times memb) + (\beta \times comp) - (\gamma \times hierdistance) - (\delta \times histdistance) \qquad (1)$$

Only certain attributes in a given domain will allow comparisons, so the authors of the domain provide a list of suitable attributes. Following M-PIRO, we consider that contrast comparisons are only interesting if the focal object is being compared with a group of at least two previous objects, since objects far apart in the hierarchy will often differ in all their attributes, but stating this would not be informative. In addition, we will only allow multiple comparisons in a single text if all the members of the group and the focal object share the value for these comparisons.

[3] The hierarchical distance is calculated as the total number of edges in the hierarchy between the FO and the CG.

## 5 An Example User Experience

### 5.1 The RCAHMS Domain

This Methodius example domain is based on a subset of the RCAHMS database of the built heritage of Scotland which was used as part of a joint project to provide a demonstrator web interface. The data includes sites in Orkney and Kinneil with a variety of site types, from Neolithic archaeology to 20th Century bridges and airfields.

A taxonomy for a fragment of the domain is shown in figure 2. Types in the ontology are shown in boxes, while actual sites are in shaded ovals.

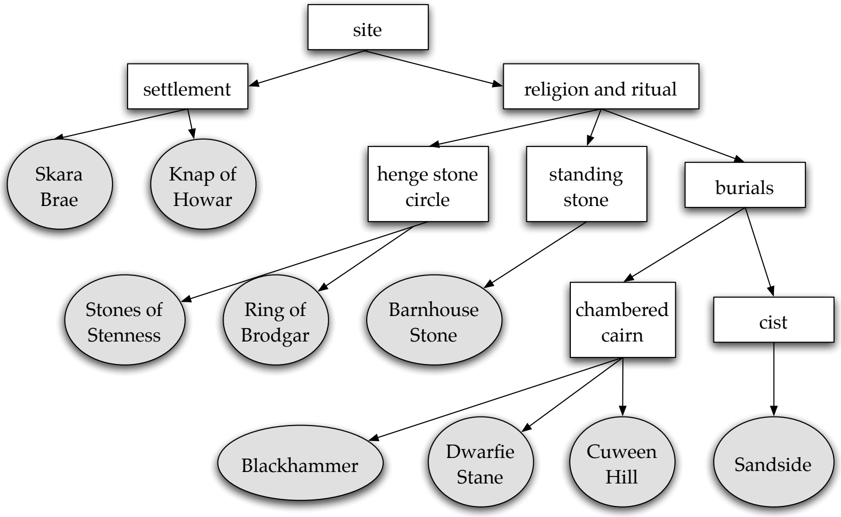


**Fig. 2.** A fragment of the RCAHMS domain taxonomy

### 5.2 User History

In the following section, we will present an example of the algorithm's results. We imagine that a user has been navigating through a web interface to the RCAHMS data like the one shown in figure 1. Table 1 shows the last 8 sites which have been viewed, and the new object which has been selected for description. The more recent an object is, the closer it is to the bottom of the table, so the selected object (FO) is the Barnhouse Stone, which is a standing stone, the previous object was the settlement Knap of Howar, the one before Skara Brae, and so on.

The attributes which have been selected for use in comparisons for the domain in this example are parish, period, and status.[4] The maximum historical distance up to which previous objects can be compared has been set to 10, so it will have no effect on this example.

| **hist** | **obj name** | **object type** | **parish** | **period** | **status** |
|---|---|---|---|---|---|
| 8 | blackhammer | chambered-cairn | rousay-and-egilsay | neolithic | guardian |
| 7 | cuween-hill | chambered-cairn | firth | neolithic | sam |
| 6 | sandside | cist | hoy-and-graemsay | early-med | N/A |
| 5 | stones-of-stenness | henge-stone-circle | stenness | neolithic | sam |
| 4 | ring-of-brodgar | henge-stone-circle | stenness | neolithic | sam |
| 3 | dwarfie-stane | chambered-cairn | hoy-and-graemsay | neolithic | sam |
| 2 | skara-brae | settlement | sandwick | neo-bronze | sam |
| 1 | knap-of-howar | settlement | papa-westray | neolithic | guardian |
| FO | **barnhouse-stone** | **standing-stone** | **stenness** | **neo-bronze** | **sam** |

**Table 1.** The most recently viewed sites and current focal object

# 6 Choosing the Best Comparison

## 6.1 Grouping Objects

The module first makes a list of all the groups of objects which have the potential to be used as comparison groups, as shown in table 2. As described above in section 3.2 , objects may be referred to by name, by type, or as “the previous X”. For example, in this table, group 5 can be referred to either as “Skara Brae and Kanp of Howar” or as “the settlements”, since they are the only two settlements in the recent browsing history.

The comparison module does not make the choice of referring expression, but it eliminates all groups for which there is no possible expression given our constraints. One an optimal comparison has been found, it will be passed to the referring expression module which will choose the most appropriate expression given the discourse history.

## 6.2 Scoring Object Groups

Once the possible comparison groups have been identified, we assign each of them a score, based on equation 1 in section 4. Groups which have no possible comparisons are discarded.

[4] The status value “sam” signifies that the site is a Scheduled Ancient Monument, considered to be of national importance, and “guardian” that it is a Guardianship Site, one which is under the full care and maintenance of a heritage organization.

| No. | Referring Expression | | | Group of Objects |
|---|---|---|---|---|
| | Name | Type | Previous | |
| 1 | N/A | site | N/A | cuween-hill, barnhouse-stone, ring-of-brodgar, dwarfie-stane, sandside, stones-of-stenness, knap-of-howar, blackhammer |
| 2 | N/A | religion-and-ritual | N/A | cuween-hill, skara-brae, ring-of-brodgar, dwarfie-stane, sandside, stones-of-stenness, blackhammer |
| 3 | N/A | burial | N/A | cuween-hill, dwarfie-stane, sandside, blackhammer |
| 4 | N/A | chambered-cairn | N/A | cuween-hill, dwarfie-stane, blackhammer |
| 5 | Yes | henge-stone-circle | N/A | ring-of-brodgar, stones-of-stenness |
| 6 | Yes | settlement | N/A | skara-brae, knap-of-howar |
| 7 | Yes | cist | N/A | sandside |
| 8 | Yes | N/A | N/A | blackhammer |
| 9 | Yes | N/A | N/A | cuween-hill |
| 10 | Yes | N/A | N/A | stones-of-stenness |
| 11 | Yes | N/A | N/A | ring-of-brodgar |
| 12 | Yes | N/A | N/A | knap-of-howar |
| 13 | Yes | N/A | N/A | dwarfie-stane |
| 14 | Yes | N/A | Yes | skara-brae |

**Table 2.** Groups of Objects which can potentially be used in comparisons

If one or more groups has possible comparisons, all the groups with the top score are processed. If the value of an attribute is the same for all members of the group, it is compared to the focal object's value for the same attribute. If the focal object has the same value, we have an illustrative comparison. If the focal object's value differs, and the size of the group is greater than one, we have a contrastive comparison. If the CG and the FO share more than one attribute value, we have a double comparison. If there is a double comparison, this will have been taken into account in the scoring process, so single comparisons for this group will be discarded, as they are less valuable. If several comparisons with the same score are generated, one will be chosen at random.

We illustrate the flexibility of the parameterized scoring system by presenting the results of the scoring algorithm with three different settings.

### 6.3 Neutral Weights

First all weights are set to 1, so we arrive at equation 2.

$$members + comparisons - hierdistance - histdistance \tag{2}$$

Table 3 shows the neutral scores for all the groups which have possible comparisons. We will use the first of these groups as an example

- 3 members: 3 (Cuween Hill, Dwarfie Stane and Blackhammer). One comparison
- 1 possible comparison: (period)
- 3 hierarchical distance between FO (standing stone) and CG (chambered cairns) (see figure 2)

- 3 historical distance - most recent in group is Dwarfie Stane (see table 1)

The equation for this group is shown in equation 3.

$$3 + 1 - 3 - 3 = -2 \tag{3}$$

| Score | Members | Comps | Hier | Hist | Reference | Group |
|---|---|---|---|---|---|---|
| -2 | 3 | 1 | 3 | 3 | TYPE:chambered-cairn | cuween-hill, dwarfie-stane, blackhammer |
| -2 | 2 | 2 | 2 | 4 | TYPE:henge-stone-circle | ring-of-brodgar, stones-of-stenness |
| -2 | 1 | 2 | 3 | 2 | NAME, PREV | skara-brae |
| -4 | 1 | 1 | 3 | 3 | NAME | dwarfie-stane |
| -4 | 1 | 2 | 3 | 4 | NAME | ring-of-brodgar |
| -5 | 1 | 2 | 3 | 5 | NAME | stones-of-stenness |
| -8 | 1 | 1 | 3 | 7 | NAME | cuween-hill |

**Table 3.** Scores with neutral weights of 1

We have three groups with the same top score, so the three best comparisons are:

- Unlike the chambered cairns, which date from the Neolithic period, Barnhouse Stone dates from the Neolithic or Bronze Age.
- Like the henge stone circles, Barnhouse Stone is located in Stenness and is listed as a Scheduled Ancient Monument.
- Like Skara Brae (*or* the previous site), Barnhouse Stone was dates from the Neolithic or Bronze Age and is listed as a Scheduled Ancient Monument.

### 6.4 Hierarchy is Less Important

However, if we decide that hierarchical distances are not so important in this domain, and set $\gamma$ to .5, yielding equation 4, the results for the top scoring comparison groups are shown in table 4.

$$members + comparisons - (.5 \times hierdistance) - histdistance \tag{4}$$

There are now two top-scoring groups, and the comparison will be chosen from:

- Unlike the chambered cairns, which date from the Neolithic period, Barnhouse Stone dates from the Neolithic or Bronze Age.
- Like Skara Brae (*or* the previous site), Barnhouse Stone dates from the Neolithic or Bronze Age and is listed as a Scheduled Ancient Monument.

| Score | Members | Comps | Hier | Hist | Reference | Group |
|---|---|---|---|---|---|---|
| -.5 | 3 | 1 | 3 | 3 | TYPE:chambered-cairn | cuween-hill, dwarfie-stane, blackhammer |
| -.5 | 1 | 2 | 3 | 2 | NAME, PREV | skara-brae |
| -1 | 2 | 2 | 2 | 4 | TYPE:henge-stone-circle | ring-of-brodgar, stones-of-stenness |

**Table 4.** Scores with hierarchy weight $\gamma$ set to .5

### 6.5 History is Less Important

If instead we decide that historical distance is not as important, and set $\delta$ to .5, giving equation 5, the results for the top scoring groups are as shown in table 5.

$$members + comparisons - hierdistance - (.5 \times histdistance) \tag{5}$$

| Score | Members | Comps | Hier | Hist | Reference | Group |
|---|---|---|---|---|---|---|
| 0 | 2 | 2 | 2 | 4 | TYPE:henge-stone-circle | ring-of-brodgar, stones-of-stenness |
| -.5 | 3 | 1 | 3 | 3 | TYPE:chambered-cairn | cuween-hill, dwarfie-stane, blackhammer |
| -1 | 1 | 2 | 3 | 2 | NAME, PREV | skara-brae |

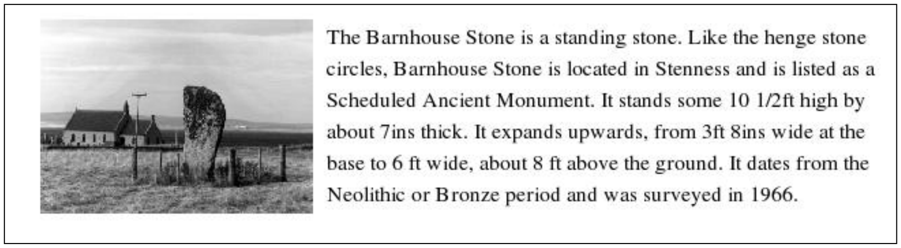
**Table 5.** Scores with history weight $\delta$ set to .5

In this case, there is only one top scoring comparison:

- Like the henge stone circles, Barnhouse Stone is located in Stenness and is listed as a Scheduled Ancient Monument.

We have shown that with minor changes to one of the four parameters, we obtain a different set of “best” comparisons. This illustrates our claim that there is no one clear best choice, and that the comparison choice must be a function of the particular circumstances in which a text is generated.

In a description system, the comparison would be part of a longer text containing other information about the object. Figure 3 shows the same description as the one in the caption of figure 1, with the addition of the comparison shown above.



The Barnhouse Stone is a standing stone. Like the henge stone circles, Barnhouse Stone is located in Stenness and is listed as a Scheduled Ancient Monument. It stands some 10 1/2ft high by about 7ins thick. It expands upwards, from 3ft 8ins wide at the base to 6 ft wide, about 8 ft above the ground. It dates from the Neolithic or Bronze period and was surveyed in 1966.

**Fig. 3.** Comparative Description of the the Barnhouse Stone

## 7 Conclusions

### 7.1 Do Comparisons Matter to Users?

A user study was carried out to evaluate the output of the M-PIRO system [1], comparing plain texts to ones which contained comparisons and also used text aggregation techniques [9]. Subjects read one set of plain and one set of enhanced texts and answered multiple choice questions and a subjective questionnaire. The results showed that they learned more and considered themselves to have learned more from the enhanced texts, and found the enhanced texts more fluent. Since aggregation and comparisons were considered together, it is not possible to separate their effects statistically, however it is likely that the comparisons provided the increase in learning, and the aggregation the increase in fluency.

### 7.2 Future Evaluation

The Methodius generation system provides a parameterizable comparison algorithm which allows flexibility in defining what constitutes the "best" comparison under the circumstances. The algorithm is based on the personalized user history, but can be customized to fit the particular domain. In a cultural heritage domain with a deep and narrow hierarchy, hierarchical distance may be less important. In a textual situation, it may be possible to provide more comparisons in a single text, since a reader has the chance to go back and check what they have read, whereas a spoken system should give less information at a time. Our algorithm aims to combine features from previous systems and domains, and provide a single solution which will fit many circumstances.

To determine the best weights to use it will be necessary to generate texts using different selections, and to perform user evaluations to determine which ones are most helpful and interesting for the users of a given system. As described above, a previous user study established that comparisons are beneficial to users in terms of learning outcomes. Natural Language Generation systems can only be evaluated in context, as they have no purpose unless embedded in a particular domain or task.

The Methodius system can be included in a system for any domain where an ontology and database of facts are available, and it is currently being used as part of the INDIGO project[5] where visitors to a museum will interact with a robot tour guide. We

[5] http://www.ics.forth.gr/indigo

plan to carry out evaluations of the system as a whole and various components individually during the course of this project.

## References

1. Isard, A., Oberlander, J., Androutsopoulos, I., Matheson, C.: Speaking the users' languages. IEEE Intelligent Systems **18**(1) (2003) 40–45 Special Issue on Advances in Natural Language Processing.
2. Reiter, E., Dale, R.: Building Natural Language Generation Systems. Cambridge University Press, Cambridge, U.K. (2000)
3. Foster, M.E.: The value of variation in automatically generated multdmodal object descriptions. PhD thesis, University of Edinburgh (forthcoming)
4. O'Donnell, M., Mellish, C., Oberlander, J., Knott, A.: ILEX: An architecture for a dynamic hypertext generation system. Natural Language Engineering **7** (2001) 225–250
5. Dale, R., Green, S.J., Milosavljevic, M., Paris, C., Verspoor, C., Williams, S.: The realities of generating natural language from databases. In: Proceedings of the 11th Australian Joint Conference on Artificial Intelligence, Brisbane (1998)
6. Milosavljevic, M.: Content selection in comparison generation. In Hoeppner, W., ed.: 6th European Workshop on Natural Language Generation (6th EWNLG). (1997) 72–81
7. Milosavljevic, M.: The Automatic Generation of Comparisons in Descriptions of Entities. PhD thesis, Department of Computing, Macquarie University (1999)
8. Melengoglou, A., Androutsopoulos, A., Calder, J., Clark, R., Dimitrimanolaki, A., Hughson, A., Isard, A., Matheson, C., Not, C., Oberlander, J., Spiliotopoulos, D., Varges, S., Xydas, G.: Generation components and documentation for prototype d4.5. Technical report, The M-PIRO Project (2002)
9. Karasimos, A., Isard, A.: Multi-lingual evaluation of a natural language generation system. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal (May 2004)