Speaking the Users' Languages

Amy Isard, Jon Oberlander, and Colin Matheson, *University of Edinburgh*
Ion Androutsopoulos, *Athens University of Economics and Business*

The authors describe a system that generates descriptions of museum objects tailored to the user. The texts presented to adults, children, and experts differ in several ways, from the choice of words used to the complexity of the sentence forms.

Suppose you're on the Web and looking at the image in Figure 1. You might ask, "What in the digital world is that?" Information about such digital objects is easy to provide to users who, for example, click on the object in their browsers. But clicking does not automatically provide information that is tailored to the individual user's abilities, interests, or preferences, or that takes account of previous exchanges between the user and the system. For that to happen, the information must be stored in a different, nontextual fashion and turned into a spoken or written message in the user's language when needed. This generation process can take into account the context and the user's linguistic and personal preferences so that it looks to the user as if the object has been stored all along as a *personalized information object*.

Usability expert Jakob Nielsen has long argued that computer systems should speak the user's language. By using personalized information objects, a system can indeed speak the user's language. Moreover, a multilingual system can speak all the users' languages. Such objects will find application in many different arenas. For instance, GPS-equipped always-on mobile devices will offer location-based services, providing users with information about nearby objects and organizations. These include both commercial and cultural products and providers: sales offers, restaurants, historical artifacts, and tourist attractions. But the attractiveness and utility of location-based services could be greatly enhanced if they trade specifically in personalized information objects. Users expect customized information, and they will expect it to be delivered using whatever presentation method makes the most sense: graphics, text, speech, or all three.

We don't need to wait for next-generation mobile devices to explore the possibilities of personalized information objects. The system we have been working on, within the M-PIRO project, has concentrated on virtual and other museum settings. A specific advantage is that museum communication professionals are well versed in the need to tailor an organization's preferred message to their visitors' expertise and information needs. The project has focused on developing language-engineering technology for personalized information objects, specifically on multilingual information delivery.

Progress on the project has required working on multilingual natural language generation from a single source, improving speech synthesis for languages other than English, generating spoken and written messages from the same source, authoring the single source through symbolic authoring techniques, and working with adaptive user modeling for personalized information presentation. The result is a state-of-the-art system that speaks the users' languages and offers highly personalized information objects.

The starting point: ILEX

ILEX (the Intelligent Labeling Explorer) was an earlier system developed at the University of Edinburgh in collaboration with the National Museums of Scotland.[1] There were two main versions of the system. The first was a Web-based virtual museum gallery. The second was a phone-based system for visitors browsing an actual physical gallery. Both used the same natural-language-generation technology at their core.

Generating language in systems such as ILEX and M-PIRO typically happens in four stages: *content selection*, *text planning*, *microplanning*, and *surface realization*. In content selection, the system selects a

1094-7167/03/$19.00 © 2003 IEEE
Published by the IEEE Computer Society

subset of the information available to it from the database, with reference to user-modeling values. Text planning involves specifying the structure of the text by ordering the chosen facts and establishing the relationships that exist between them. In the microplanning stage, the system carries out two main tasks, taking into account current and prior context. On the one hand, the system chooses abstract specifications of verb and noun phrases (processes known as *lexicalization* and *referring expression generation*). On the other hand, it combines sentences together into more complex structures (a process known as *aggregation*). Finally, the structures created by the microplanning are turned into actual text.

Providing personalization

When interacting with the ILEX Web Demonstrator, visitors start from an index page of thumbnail images. Clicking an image causes the system to generate a description of the selected object and display it with a full-size image. There is no separate introductory page because background information is incorporated into the personalized descriptions generated on demand. Although a description contains links to recommended objects, the visitor can return to the visual index and choose a new object at any time.

We designed ILEX's descriptions to convey accurate information that is both important and interesting. Information is important when it helps educate the visitor more broadly; it is interesting when it holds the visitor's attention. To help meet these criteria, ILEX provided facilities for simple user-types, a discourse history, and its own agenda of communicative goals. The user has the freedom to explore any personalized information object at any time, potentially making completely unanticipated jumps. However, the descriptions produced are constrained by the system's own agenda of educational goals, which it attempts to achieve whenever the opportunity arises.

To this core functionality, the speech-enabled version added a link between ILEX and the Festival speech synthesis system.[2] Instead of simply sending text to be spoken, ILEX sends text marked up with information about what is being talked about and how it is related to what it has previously sent. This technique led to some improvement in the quality of the synthesized speech.[3] More generally, we found that users learned just as much from personalized information objects as they did from impersonal information—even though users clicked on 50-percent fewer links.[4]

Drawbacks

To a certain extent, ILEX was successful, but the system had several drawbacks. For example, it was expensive to provide knowledge-base information from which ILEX could generate text. We imported some information from the museum's own information-management system. But we captured much of it by interviewing a curator and then hand-coding taxonomic information and other assertions. From the start, we incorporated a workaround into ILEX that let us type text strings literally rather than in terms of knowledge-base objects. However, every time we used this facility, it made it impossible to present information in a language other than the original input language.

In addition, ILEX's linguistic resources were based on a grammar developed initially for English. Small-scale extensions let us try a version for Spanish, but it became clear that the system did not directly support multiple languages. The main problem was that linguistic and nonlinguistic domain information was interwoven through various parts of the system.

Also, ILEX successfully carried only one particular kind of user modeling. The content included in a personalized information object was influenced by what the user had seen previously in that session. But the system retained no persistent information about a user. So a second visit would not be influenced by what had happened in the first. Furthermore, the precise way in which the system referred to objects was influenced, within a session, only by discourse history. This means, for instance, that the system could use a pronoun, such as "it," after an object had been previously mentioned. But otherwise, there was no sense in which the form of descriptions was influenced by user type. An expert would not see more technical terminology than a member of the general public. A child would not read simpler or shorter sentences than an adult. In fact, partly for this reason, the version of ILEX used for user evaluation did not even activate these differing user stereotypes.

Finally, ILEX suffered from a general problem that often arises in research prototypes: The system was less modular than is desirable. We have already pointed out that language-specific constraints appeared in many parts of the system, which meant that processes and linguistic resources were not cleanly separated. We also noted that user modeling affects some, but not all, levels of language planning. But another problem relating to modularity is that unless linguistic resources are very broad in their coverage, they tend to be tied to the domain being modeled, which means that the domain knowledge base cannot be replaced easily by another without requiring changes in the various places where linguistic information is encoded.

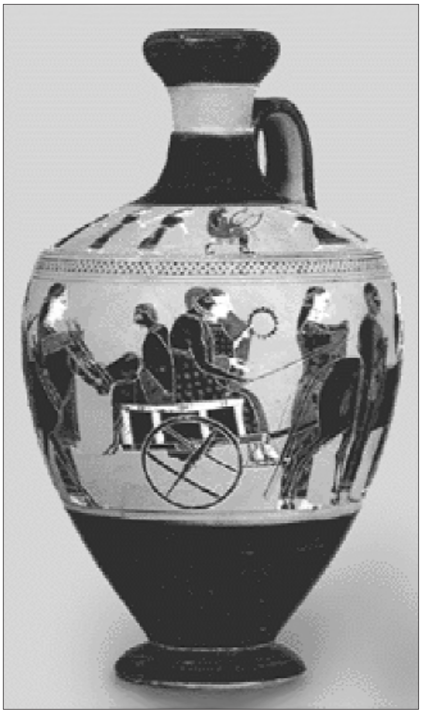

Figure 1. What in the digital world is that? Clicking on an object in a Web browser does not typically provide information tailored to the user's interests.

The state of the art: M-PIRO

While ILEX served up personalized information objects, there was substantial room for improvement. With M-PIRO, we made progress on several fronts. One clear advance is that we developed an authoring tool (developed at NCSR "Demokritos") to help museum creators create and edit the domain-specific knowledge base and linguistic resources. Figure 2 shows a view of the current museum domain.

Domain authoring

The knowledge base contains information

about entities and relationships between entities. Entities can be both abstract (such as historical periods or painting styles) and concrete (such as a particular amphora in the collection). The authoring task happens in two stages: domain authoring and exhibit authoring. The more work that can be done in the first stage, the less work is needed later.

One of the first tasks in domain authoring involves setting up a hierarchy of entity types. At the top of the hierarchy are basic types, such as *location* and *exhibit*. Next come entity types such as *museum*, a subtype of *location*, and *vessel*, a subtype of *exhibit*. There can be more than one level of entity types. For example, vessel contains the subtypes *amphora* and *lekythos*. To make the authoring tool easier to use, we built a single-inheritance hierarchy; each entity type can have only one direct supertype, and each entity has only one direct entity type parent.

We express relationships between entities using fields. The domain author can define fields for each entity type, which are then inherited by all entity types below it in the hierarchy. For example, *exhibit* has fields *current-location* and *exhibit-depicts*, which then also apply to *vessel*, *amphora*, and *lekythos*. Each field-filler must be specified as belonging to a particular entity type or built-in data type (such as a string or date). For example, the *current-location* field must be filled by an entity of type *museum* and the *exhibit-depicts* field must be filled by a text string.

Each field also has associated information that specifies how the relationship it represents can be expressed as a sentence. The specifications for these are known as *microplanning expressions*. The domain authors can create one or more of these expressions, either as a clause plan or as a template. If they choose a clause, they can then select a verb using a pull-down menu of the verbs defined in the domain-dependent lexicon, along with optional prepositions and modifiers. If they choose a template, they can build the expression using strings and references to the two entities whose relationship is expressed by the field. Templates are available so that the domain authors can choose to have more direct control over the output from the system without going so far as to allow them to specify completely fixed text strings.

The next task for the domain authors is to populate the domain-dependent lexicon, which contains entries for nouns and verbs. Function words, such as articles and prepositions, are domain independent and are stored as a separate resource that separate domains can share. Each noun is associated with an entity type, which lets the generation engine use a particular noun when referring to an entity of the appropriate type. For example, the noun *exhibit-noun* is associated with the basic type *exhibit* so that the system can generate the noun phrase *this exhibit* in English, *αυτό το έκθεμα* in Greek, or *questo reperto* in Italian. Each entity type also inherits the nouns associated with its supertypes. The domain author associates at least one noun with each entity type, adding new nouns to the lexicon as necessary.

The domain author also adds domain-dependent verbs for use in microplanning expressions. The authoring tool supports entering noun and verb forms. For instance, in Greek you can generate noun forms on the basis of the nominative singular and plural forms. You can also inspect and amend the automatic results if necessary. Figure 3 shows the Greek lexicon being authored, with the various forms of lekythos.

> We designed the authoring tool to be used by people, such as museum curators, who have no experience in language technology.

Exhibit authoring

Compared with domain authoring, exhibit authoring is the much simpler process of defining specific instances of entities and filling their fields with the appropriate information. The authoring tool has a facility for previewing the output from the generation system on the basis of the current state of the database. Exhibit or domain authors can use this facility to check that the information they have entered results in appropriate texts.

For example, the domain author could define

- a basic type *location*
- an entity type *museum* that is a subtype of *location*
- a basic type *exhibit* that has a field *current-location*, which must be filled by an entity of type *museum*
- a microplanning expression for each language for the field *current location*, which in English specifies the use of *locate-verb* in the present tense and passive voice, with the preposition *in*
- an entity type *vessel* that is a subtype of *exhibit*
- an entity type *lekythos* that is a subtype of *vessel*

The exhibit author can then create an entity *new-york-metropolitan-art* of type *museum* with a name field for each language and an entity *exhibit* of type *lekythos*, whose current-location field is filled with the *new-york metropolitan-art* entity.

In Figure 2, a designer is selecting *new-york-metropolitan-art* for *exhibit15* to fill the field *current-location*. The system could use this information to generate the following sentences:

- This exhibit is a lekythos. It is located in the New York Metropolitan Museum of Art.
- Questo reperto è una lekythos. Si trova nel Metropolitan Museum of Art di New York.
- Αυτό το έκθεμα είναι μια λήκυθος. Βρίσκεται στο Μητροπολιτικό Μου–σείο της Νέας Υόρκης.

We designed the authoring tool to be used by people, such as museum curators, who have no experience in language technology. More training would be necessary for the domain-authoring stage, but we think it would be possible to train a few museum staff who would be responsible for setting up the new domain and making later modifications to the structure as necessary. Once you complete the initial setup, a large number of less-advanced users can do the exhibit authoring. Initial experiences suggest that it might still be necessary to have a trained linguist check the microplanning expressions to make sure that they have been correctly specified.

You can also use the authoring tool to create user types and attach user-modeling information to fields and microplanning expressions. Once you complete the authoring, the system exports the user-modeling information into the personalization server, and exports the lexicon and domain data as XML files, which can then be loaded into the generation system.

Three languages

M-PIRO can currently generate text in three languages: English, Greek, and Italian. The grammar resources are language-independent as much as possible.[5] This means that we exploit the similarities between the grammars of different languages, thereby easing resource development and maintenance. We based the central grammatical representations on the English resources from the ILEX system, which are in turn based on the WAG system.[6] We constructed the Greek and Italian grammars by starting with the ILEX English grammar and a grammar of Spanish, and then incrementally modifying the elements where the languages differ.

The domain resources are also now largely language-independent. On the one hand, the three languages share the hierarchy of entity types and their fields. On the other hand, the lexicons, microplanning expressions, and text string field-fillers are language-specific.

M-PIRO also provides high-quality speech output in the three available languages. The Festival speech synthesis system provides English and Italian speech, and the Demosthenes system (developed at the University of Athens) provides the Greek. We can configure the generation system so that its output is in the form of an XML-based speech markup language, which will allow the synthesizer to make use of information including parts of speech, sentence boundaries, rhetorical relations, and other distinctions that can be used to improve the prosody of the speech output.

User modeling

We store the user-modeling information separately from the domain and linguistic resources in the personalization server (developed at NCSR "Demokritos"), a database that is interrogated by the system during generation and dynamically updated after each text is produced. This technique allows user-modeling information to persist over time and means that if a single user accesses the database from multiple locations (at a museum and then later from home via a Web page, for instance), the system always has access to the user's personal profile and information on the history of the user's interactions with the collection.

As we mentioned, the domain authors can define one or more user types. In the current M-PIRO system, there are children, adults, and experts. Each entity type field has values for interest, importance, and repetitions

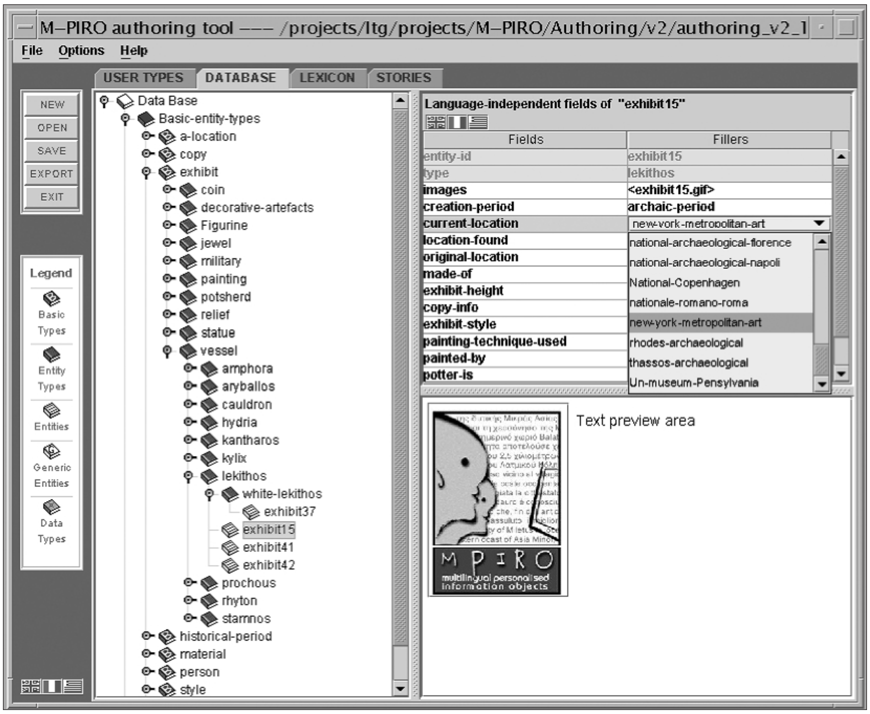

Figure 2. Domain authoring. A museum curator uses the authoring tool to specify that exhibit 15, a lekythos, is currently located in the New York Metropolitan Museum of Art. On the left you can see the hierarchy of objects in this museum domain.

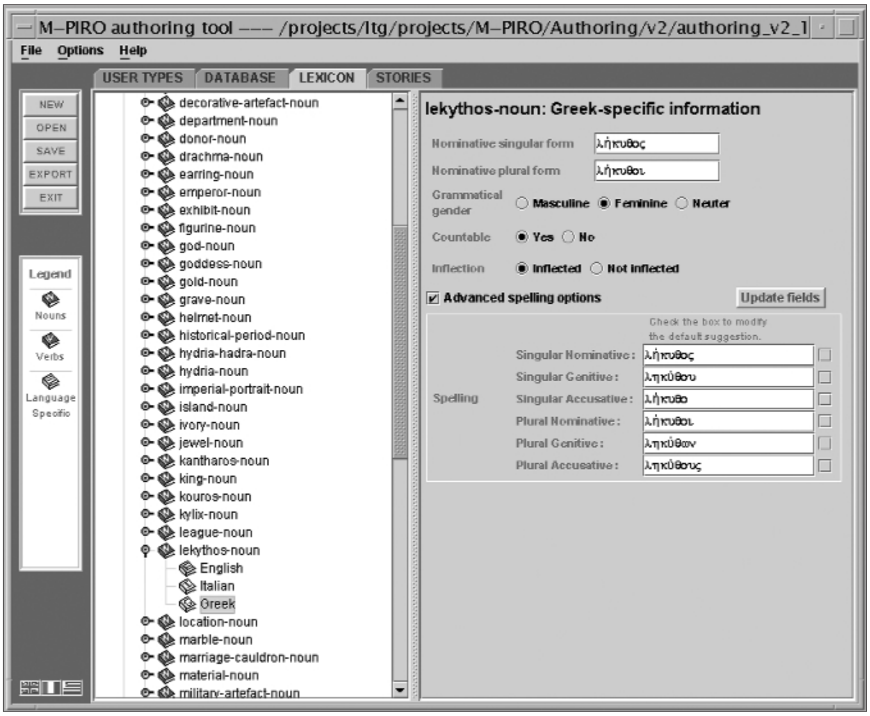

Figure 3. Lexicon authoring. On the left is a list of nouns that are used in descriptions of the exhibits in this collection. On the right, an author specifies the different grammatical forms of the Greek noun λήκυθος (lekythos).

for each user type assigned to it during the authoring process. By default, entities further down the hierarchy inherit these values, but the values for a field and a particular entity (a fact) can be redefined.

The interest score quantifies how likely a visitor of a particular type is to find the fact interesting. For example, experts might be interested in a list of references to published articles that discuss the selected exhibit, while children might be more interested by a description of the purpose of the object. The importance score reflects how important the museum curator thinks that the fact is for each user type.

The system uses the *repetitions* value to calculate the assimilation score and assimilation rate for each user type. The assimilation rate is set to 1/repetitions, which means that each time the system presents a fact to the user, the assimilation score is incremented by the assimilation rate. For example, we set the repetitions value for *creation-period* to 1 for experts but 2 for adults and children. This means that the system will present a fact about the creation period of an entity to an adult or child user twice before the system considers the user to have fully assimilated the information. Each time the system generates a description, it will update the user's assimilation scores. For each exhibit, the database typically contains more information than can be expressed in a description of reasonable length. The system attempts to convey only unassimilated facts that have not been expressed in the past. If there are too many of these, it chooses those high in interest and educational value.

During domain authoring, the system sets values for each user type that specify the maximum number of facts that should appear in one description and the maximum number of facts that should appear in one sentence. The system can use these values to vary the complexity of the generated text. M-PIRO contains an aggregation module that uses techniques such as simple conjunction, relative clauses, and syntactic embedding to join together single facts.[7] If the system aggregates many facts, the sentence structure can become complex, so the maximum facts per sentence value for children can be set lower than for adults or experts.

Using the authoring tool, the designers can supply multiple microplanning expressions for each field. Each one can have different interest and importance values assigned to it. This capability lets us vary the way in which the system expresses the same fact and to use different ways of expressing the same fact for different user types.

Modularity

M-PIRO's system architecture is significantly more modular than that of its predecessor ILEX. In particular, the linguistic resources, database, and user-modeling subsystems are now separate from the systems that perform the natural language generation and speech synthesis.

Of course, it is not possible to move to a new application domain without specifying both what will be talked about and what vocabulary will be used when talking about it. But there is now a clean separation between what is specific to a domain and what is independent of it. The authoring tool makes it clear to the authors when they have to provide new linguistic information, as well as domain information.

> If we want our intelligent virtual guide to be able to talk about more than a few dozen exhibits, we might have to devote a serious amount of time to data entry.

The way ahead

Can you use M-PIRO's technology to liven up your Web site? In many cases, the answer to this question is yes. And the applications of M-PIRO's technology are not limited to generating personalized Web pages. We have already mentioned the possibility of using M-PIRO's software in next-generation mobile devices or to drive animated characters that will act as virtual salespeople or guides. Our colleagues at the Foundation of Hellenic World in Athens are already working on a new M-PIRO prototype that will embed the project's technology in their immersive virtual reality system. This will let their visitors receive personalized information in spoken form as they wander in FHW's virtually reconstructed archaeological sites.

However, there are limitations to the technology. First, as with its predecessor ILEX, M-PIRO focuses on object descriptions. You cannot use the project's software to describe processes like those that would appear in manuals. Other natural-language-generation researchers have explored this problem.[8]

M-PIRO's authoring tool is a big step forward from ILEX. Authors no longer need to be programmers or natural-language-processing experts. They interact with a much higher-level tool that lets them manipulate information and control how this information will be rendered in natural language. But it could well be that a significant amount of information still needs to be entered for each new exhibit during the exhibit authoring phase. If we want our intelligent virtual guide to be able to talk about more than a few dozen exhibits, we might have to devote a serious amount of time to data entry.

Another improvement of M-PIRO compared to ILEX is that it provides, at the end of the domain-authoring phase, a clear description of the information that the generation engine needs to know for each exhibit in the form of an XML DTD (document type definition). The generation engine is indifferent about where this information originates, provided that it conforms to the DTD. It is therefore possible to establish XSLT (Extensible Stylesheet Language Transformations) mappings between the generation engine and existing external databases to save the authors from having to reenter information that already exists in other formats.

A multilingual natural-language-generation system becomes useful when it supports many languages. M-PIRO has paved the way by providing a generation architecture that can demonstrably accommodate multiple languages. But the current language set—English, Greek, and Italian—would have to be extended for the software to be viable in multilingual markets such as Europe. A future open-source version of M-PIRO's software might be a good way to invite specialists from other countries to develop the necessary resources for their languages.

Museum curators must be sure that the generated descriptions are of high quality, in terms of content and language. M-PIRO's authoring tool lets curators preview generated text for any type of user and see the effect of the various parameters on the resulting text. But, as an evaluation at FHW has already made clear, this process would be much simpler if the authors could click on problematic pieces of generated text (for instance, on a word or phrase) to be taken automatically to the corresponding database entries or language resources (for instance, a lexicon entry or microplan) that are responsible for the corresponding text. Being able to do this calls for a more active preview form that moves towards the What You See Is What You Meant approach,[9] whereby the author interacts with an initially simplistic piece of text that becomes more elaborate as more information is added.

We hope to see question-answering capabilities added to M-PIRO's technology, like the ones that let users submit natural language questions to document collections. In our case, however, the solution would be a dynamically generated text rather than a snippet of an existing document. Approaching an exhibit or a particular location might be a good indication that you want to be given information about it. But in many cases, it is easier to state what you want to be told in the same way that you can tell a salesperson what you are looking for.

We started out by supposing that you're on the Web and looking at the image in Figure 1. The question was, "What in the digital world is that?" Because M-PIRO speaks the users' languages, its answer in English to a nonexpert adult seeing the object for the first time is, "This exhibit is a lekythos, created during the archaic period. It dates from circa 500 BC. It was painted by Amasis with the red figure technique and it originates from Attica." ■

The Authors

Amy Isard is a research associate in the School of Informatics at the University of Edinburgh. Her research interests include natural language generation and annotation of dialogue corpora. She received an MSc in artificial intelligence from the University of Edinburgh. Contact her at the School of Informatics, Univ. of Edinburgh, 2 Buccleuch Pl., Edinburgh EH8 9LW, UK; amy.isard@ed.ac.uk.

Jon Oberlander is a reader in cognitive science at the University of Edinburgh's School of Informatics. His research interests include automatic discourse generation and diagrammatic reasoning and communication. He received a PhD in cognitive science from the University of Edinburgh. Contact him at the School of Informatics, Univ. of Edinburgh, 2 Buccleuch Pl., Edinburgh EH8 9LW, UK; j.oberlander@ed.ac.uk.

Ion Androutsopoulos is a lecturer in the Department of Informatics, Athens University of Economics and Business. His research interests include question answering, natural language generation, document filtering, and information extraction. He received a PhD in artificial intelligence from the University of Edinburgh. Contact him at Dept. of Informatics, Athens Univ. of Economics and Business, Patission 76, GR-104 34 Athens, Greece; ion@aueb.gr.

Colin Matheson is a research fellow at the University of Edinburgh. His research interests include computational linguistics, especially theoretical and applied models of dialogue. He received a PhD in cognitive science from the University of Edinburgh. Contact him at the School of Informatics, Univ. of Edinburgh, 2 Buccleuch Pl., Edinburgh EH8 9LW, UK; colin.matheson@ed.ac.uk.

Acknowledgments

M-PIRO is a project of the Information Societies Programme of the European Union. The project's consortium consists of the University of Edinburgh (UK), ITC-irst (Italy), NCSR "Demokritos" (Greece), the University of Athens (Greece), the Foundation of the Hellenic World (Greece), and System Simulation Ltd. (UK). The work by Ion Androutsopoulos on this project was carried out at the Institute of Informatics and Telecommunications, NCSR "Demokritos," Greece

References

1. M. O'Donnell et al., "ILEX: An Architecture for a Dynamic Hypertext Generation System," *Natural Language Eng.*, vol. 7, 2000, pp. 225–250.
2. P.A. Taylor, A.W. Black, and R.J. Caley, "The Architecture of the Festival Speech Synthesis System," *Proc. 3rd Int'l ESCA Workshop Speech Synthesis*, 1998, pp. 147–151.
3. J. Hitzeman et al., "On the Use of Automatically Generated Discourse-Level Information in a Concept-to-Speech Synthesis System," *Proc. 5th Int'l Conf. Spoken Language Processing*, IEEE Press, Piscataway, N.J., 1998, pp. 2763–2766.
4. R. Cox, M. O'Donnell, and J. Oberlander, "Dynamic Versus Static Hypermedia in Museum Education: An Evaluation of ILEX, the Intelligent Labeling Explorer," *Proc. 9th Int'l Conf. Artificial Intelligence and Education*, IOS Press, Amsterdam, 1999, pp. 181–188.
5. J. Bateman et al., "The Reuse of Linguistic Resources across Languages in Multilingual Generation Components," *Proc. 12th Int'l Joint Conf. Artificial Intelligence*, AAAI Press, Menlo Park, Calif., 1991, pp. 966–971.
6. M. O'Donnell, *Sentence Analysis and Generation: A Systemic Perspective*, PhD dissertation, Linguistics Dept., Univ. of Sydney, 1994.
7. A.C. Melengoglou, *Multilingual Aggregation in the M-PIRO System*, MSc Thesis, School of Informatics, Univ. of Edinburgh, 2002.
8. C. Paris et al., "A Support Tool for Writing Multilingual Instructions," *Proc. 14th Int'l Joint Conf. Artificial Intelligence*, AAAI Press, Menlo Park, Calif., 1995, pp. 1398–1404.
9. R. Power and D. Scott, "Multilingual Authoring Using Feedback Texts," *Proc. 17th Int'l Conf. Computational Linguistics*, 1998, pp. 1053–1059.