First Year Research Report: FNK PhD-Stipendiumsprojekt
University of Bremen

Juan Rafael Zamorano Mansilla
1st January 2003

1. Overview

According to the plan included in the proposal of this research project, the first year would be devoted to two main tasks:

a) an in-depth exploration of all the operations and capabilities of the KPML natural language generator, accompanied by an initial evaluation of their possible application in language teaching;

b) a study of the groups that might benefit from the use of multilingual generators.

These two tasks have now been satisfactorily completed. However, although the second task demanded less time than anticipated in the research plan (eight months), this was compensated for by the fact that two new tasks not included, or under-estimated, in the initial plan arose. These were developed in the time remaining after the potential groups of users of generators for learning purposes were defined—which means that the overall project is still running on schedule.

The new tasks were:

a) **the creation or improvement of linguistic resources for Spanish, English and German sufficient to enable the design of exercises for students.**

The reason why this task had not been anticipated in the plan is that it only became evident that some of the linguistic resources were not comprehensive enough to design a sufficiently ample variety of activities when actually trying to implement language exercises in KPML. This is particularly true of the Spanish grammar, that has required a considerable effort to obtain resources comprehensive enough. The grammar of English, on the other hand, has required additions at a smaller scale, while the grammar of German is still to be updated.

b) **a description of the additions the programme would require if used for teaching purposes.**

As stated in the proposal, text generators are programmes designed to produce an output text in a natural language, not to assist in the process of language learning. If they are to be used by teachers and learners of languages, it is evident that their characteristics must be adapted to the demands imposed by their new function. In the design of these additions to the programme two main factors have been taken into account: the range of language exercises that could be developed for KPML and the needs of its potential users. A more detailed description of the additions is given in section 5, but in general we can say that

they respond to four basic requirements imposed by language learning. These requirements are:

- the provision of an easier way of manipulating the input for generation;
- the possibility of carrying out certain operations on the linguistic forms that were unnecessary for generation;
- a specification of the quantity and quality of the linguistic information that should be presented to the user depending on his/her needs, as well as how this should be displayed; and
- a simplification of the interface to prevent the user from being confused by an excess of options that are not relevant to language learning.

The research conducted in this first year has produced the following results:

a) a demonstration at the Linguistics Colloquium (Institut für Algemeine und Angewandte Sprachwissenschaft, IAAS, FB10, Universität Bremen) held in June 2002, where various language activities for English and Spanish language learning/teaching implemented with KPML were first shown;

b) a generation bank for the Spanish language has been made publicly available on the website http://www.fb10.uni-bremen.de/anglistik/langpro/kpml/genbank/generation-bank.html This generation bank contains around 140 semantic specification and surface sentence pairs that illustrate phenomena from different areas of the grammar of Spanish; a listing of the sentences covered is attached as an appendix. At present this linguistic resource represents the largest computational generation grammar for Spanish available worldwide.

c) An article is currently under preparation in collaboration with the research project supervisor, Prof Dr. John Bateman, containing some examples of language activities for English and Spanish implemented with KPML. The article is to be submitted to an appropriate journal dealing with the application of computers to language learning.

More detailed information about the progress made concerning each of the four tasks mentioned above is given in the following sections.

2. Evaluation of generators as teaching tools.

After examination of previous applications of computers in language learning, the exercises found in language courses and the characteristics of KPML, five main lines of investigation have been established, corresponding to five areas in which generators could have worthwhile exploratory applications: the use of generators to provide grammar exercises at or below the sentence level, the use of generators as a tool for improving writing skills, the use of generators as a searchable corpus, the use of generators as tools for working with texts and the use of generators as tools for grammatical analysis.

2.1.   *The use of generators to provide grammar exercises at or below the sentence level*

Under the category of exercises at or below the sentence level I have included activities that usually consist of sentences or short texts that the student must complete or change in a certain way to practise a pattern of the language (for a good collection of exercises that belong in this category and are easily computerised, see Appendix A in Kenning & Kenning 1984). It

is a type of activity characteristic of the first stages of the learning process, that is, addressed to beginner students, and it deals with basic morphological and syntactic phenomena in which the learner must achieve an acceptable degree of automation through practice. Appendix A shows some examples of this type of activity for English and Spanish adapted to KPML. The exercises have been taken from Sánchez Lobato & García Fernández (1999) for Spanish and Murphy (2000a) and Murphy (2000b) for English.

Exercises like these are easily provided by generators, since we only have to select the sentences (or create our own bank of sentences, if we prefer) that are relevant for our purposes and make the desired changes (hide some words, give only the base form of one of the constituents, change certain properties of the sentences such as polarity, mood, voice). A full generation gives the right solution, which can be presented directly to the learner or can be used by the programme to offer feedback after comparison with the learner's answer.

A first evaluation before the actual tests take place says that the most valuable property of generators when applied to these exercises is re-usability. Any sentence can be used for almost any type of exercise, which means that exercises are not specifically built to practise exclusively a particular point of the grammar. As an example, let us say that if the teacher decides to work with the plural of nouns in English, it is not necessary to have specific sentences for that purpose; it suffices to select a number of sentences that lend themselves to manipulations with the number of nouns. In this aspect, generators share with authoring languages such as Pilot or EnBasic (cf. Underwood 1984:81-97) the feature that the teacher has a number of tools at his/her disposal to carry out a wide range of operations to design the activities, instead of having a number of fixed activities with sentences that can only be exploited in a certain way.

Another interesting property of generators is that it is extremely easy to enlarge the number of sentences available, with the advantage that the programme always offers the solution to the newly created sentences. The enlargement is made through changes in the input data required for generation and in principle can yield an unlimited number of new sentences.

A third significant advantage of generators is in the quality of the feedback they can potentially provide. Since generators contain the necessary linguistic resources, they can compare the learner's answer not just with the right answer, but with a whole range of grammatical possibilities derived from the grammar. Using a concrete example, if we are working with the conjugation of the verb, when the programme finds a mismatch between what the learner types in and the right solution, the programme can then generate all the forms of the verb to check if the mistake is the result of a confusion in the conjugation. This is very different from what we often find in other programmes (cf. Witton 1992, Robinson 1989), in which the programmer must predict the possible answers from the learner and include them in the programme. In contrast generators have in theory the capacity to do this automatically. In this way KPML should be able to recognise mistakes in all kinds of morphological variation (case, number, gender, conjugation), word-order and syntactic constructions (questions, negations, comparisons, etc). This information could then be used to offer more relevant and precise comments on learners' answers.

2.2. *The use of generators as a tool for improving writing skills*

Another area in which generators could have an application is writing. Actually the use of generators as a tool for developing writing skills is often found in the short history of CALL. From SHRDLU (Winograd 1972) and ELIZA (Weizenbaum 1976), programmes that simulated natural interaction, to language-teaching software such as LUCY (Stevens 1986), JOHN AND MARY (Higgins and Johns 1984), TENSELAND (Higgins and Johns 1984) and PARRY (Kenning and Kenning 1983). Hope et al. do not mention the programme they use,

but they offer a list of activities that can be performed with simple generators (Hope-Taylor-Pusack 1984). A review of the first attempts shows, however, that some of these early generators were not sophisticated enough to be used as real assistants in the process of creating a text. Rather their output text was treated as an object of discussion and evaluation by the students, who had then to alter it to obtain a more natural text or simply discuss the contexts in which such texts could make sense. KPML is in my opinion sophisticated enough for its output texts to be regarded as models of well-structured and well-constructed samples of the target language, and consequently serve as assistants for the development of writing skills.

This potential application has been divided, depending on the complexity of the text to be produced, into basic and advanced writing.

Basic writing concerns only short texts, usually consisting of a single sentence or a clause complex at most, and comprises the following types of exercise:

a) word-order. The student is presented with sentences in which words have been shuffled, and s/he must re-organise the text.
b) writing with hints. The student receives a number of words less than the total number of words that constitute the sentences, and must build on these hints so as to produce a good sentence.
c) Re-writing. The student is forced to re-express the content of a sentence through a second sentence that typically exhibits a completely different structure.
d) Complete with your own ideas. The student must complete a sentence with his/her own ideas, although the context forces him/her to use a specific structure.
e) Brief descriptions of pictures. The student must describe the content of a picture, or rather, produce sentences that are based on information from the picture. The sentences are not related to each other, and so do not constitute a longer text. The result is just a collection of sentences describing different aspects of the visual information.

This type of short text is generated by KPML from the same type of input as for ordinary sentences (SPL's), and the way many clauses combine to form a clause complex follows the framework established by the Rhetorical Structure Theory (cf. Mann and Thompson 1987; also the thesis proposal and research plan). Two additions, however, will be necessary if generators are to be used by learners of foreign languages:

a) a more intuitive and user-friendly interface for the creation of SPL's, as well as a tool that will deal with the relations between clauses. Right now supra-clause relations are integrated in the SPL's, but it would be convenient for teaching purposes to have them separate.
b) a tool that performs the kind of operations required for the activities described above, such as changing the order of words randomly, showing the base form of each word, hiding some of the words, etc.

Advanced writing refers to a more elaborate type of compositions in which sentences are connected in a form that gives way to a cohesive and coherent text. Whereas KPML itself does not support this type of generation, the necessary components—text planner and discourse resources—have been produced for KPML in earlier projects. The potential of adopting this more advanced mode of generation to the current task is well worth investigating. This means, however, that some time of the next two years should be put aside for the task of refining this component and probably creating or improving the resources for

Spanish, English and/or German, as well as designing an easier interface for non-experts that will guide in the process of text generation.

2.3. The use of generators as a searchable corpus

The usefulness of corpora in second language learning has been pointed out by many authors (cf. Biber, Conrad and Rippen 1998, Johns 1993, McEnery, Wilson and Baker 1997, Granger 1998), and there are presently some projects working on the creation of computational corpora for their pedagogical exploitation such as the Chemnitz Internet Grammar of English (Schmied 1999). A generator like KPML can in principle be used in the same way, here the corpus is the collection of sentences stored in the generation banks for each language. The main difference with respect to an ordinary corpus is that searches can be done according to one criterion not usually available in other corpora: namely, the semantics of the sentence. Since KPML is a semantics-oriented generator, it is possible to make a search based on the semantic features that were activated in the process of generation. This trait can be of particular interest in cross-linguistic comparisons.

Two needs arise from the use of KPML as a corpus:
  a) it will be necessary to create a tool that permits the performance of searches of different kinds: by form, by syntactic pattern, by semantics, etc.
  b) it is fundamental to have a good number of examples to form a reliable corpus. Unfortunately the process of incorporating examples in the generation banks is a slow one, even for expert users. It is to be hoped, however, that the interface for the creation of SPL's currently in preparation will speed up the process. At any rate a Spanish corpus containing around 70 examples with the verbs *ser* and *estar* is in the process of being created, to be followed by an English corpus probably illustrating the use of spatial prepositions. This may, therefore, also provide potentially useful input for the newly started Sonderforschungsbereich in Spatial Reasoning and Communication at the Universities of Bremen and Freiburg.

2.4. The use of generators as tools for working with texts

In the history of CALL there has been a particular group of programmes with certain success that based their activities on texts that were usually fed into the programme by the user. Probably the most popular one was Storyboard, written in 1982 by John Higgins but based on two previous programmes designed by Tim Johns, Masker and Textbag (Higgins and Johns 1984). Later the programme would become CopyWrite, an updated version that appeared in 1984. By the end of the 80's the idea was popular enough to produce new versions for different languages and with different capabilities. Among these we can include Developing Tray, TextPlay, Storyline, Quartext, Storycorner, Memory and GuessText (Levy, 1997). The activities included things such as hiding some (or all) words, jumbling the paragraphs or the sentences, hiding some letters, changing the gender of the characters of the text, etc.

Any of these activities could be performed with KPML, provided that a new tool is added to the programme which implements the desired changes in the text. Of course the texts would be those generated by the Text Planner, and they could be chosen from a generation bank or generated *ad hoc* by the user. After revising some text-based activities found in ordinary books and the activities created by programmes such as Hypertext, it has been decided that such a tool should be able to carry out the following operations on a text:

  • hide some letters (or all of them)

- hide words at random
- hide words according to morphological criteria (nouns, prepositions, etc)
- hide words according to functional criteria (subject, direct object, etc)
- hide groups and phrases
- hide punctuation (and accents in some languages)
- hide morphological endings (endings for gender, number, case, person, tense, etc)
- hide sentences
- change participants (which triggers changes in the pronouns referring to them, and, depending on the language, changes in the adjectives, verbs and nouns that establish agreement with them)

Once more generators exhibit an important advantage over other programmes that work with texts: since their texts are the result a traversing the grammatical resources, they simply contain more information about its structure and constituents. More specifically, when hiding any of the elements listed above, KPML could give semantic hints about the element that is missing. For example, if we remove the verb *is* in a sentence, the programme could show the message "relational process" or "finite" or "verb". If we remove an element like *The dog*, it could give hints such as "subject", "actor" or "nominal group".

2.5. The use of generators as tools for grammatical analysis

Outside the area of second language learning there is a possible application of generators in more specialised studies of language. It is the grammatical analysis of linguistic units that constitutes part of the programme of certain university courses.

Below university level this application could also be useful in the Spanish education system, in which a great emphasis is given to this type of analysis both in primary and secondary education. The grammatical analysis at this level is based on the concepts and terminology of traditional grammar and is carried out in the following way: for each constituent in the sentence the student must provide two pieces of information: the formal classification of the constituent (noun phrase, verb phrase, prepositional phrase, noun, adjective, etc); and the function it realises in the sentence or phrase (subject, direct object, head of the phrase, determiner, etc).

KPML already contains an option that displays a tree-analysis of the constituents of the clause and their structure, which makes it suitable for this purpose. But the use of generators for grammatical analysis involves the tailoring of the information displayed by the programme to the user's needs. Needless to say the level of exhaustiveness demanded is different in the case of students of primary school and university students of linguistics. This adaptation will be carried out by a special component that will be added to the programme.

A further difficulty is that generators do not analyse text, they produce it. As a consequence it is necessary to generate a sentence before the grammatical analysis can be displayed. If they want to, the users can generate their own sentences making use of the simple interface to be created and described in section 2.1. Alternatively the user may pick one of the sentences stored in the generation banks.

3. Potential groups of users and needs

The students who might be interested in the use of generators for didactic purposes within the scope of the current project have been divided into three groups:

a) learners of a foreign language
b) university students of linguistics
c) Spanish students of primary and secondary school

Some common features, however are shared by the three groups. One of them is that they need simpler menus than the ones currently displayed by KPML. The main drawback for the non-expert is that the programme actually has too many options, far more than required by the users described here, and this may cause confusion. Another obstacle is that some operations such as the modification of SPL's cannot be done from inside the programme, but either require the modification of an external file that is subsequently loaded after being saved or familiarity with the editing programme Emacs, with which KPML can communicate directly. The utility of previous tools in this area (such as the SPL authoring tool 'SPLAT') needs to be evaluated. Finally it is very likely that the users described here might demand a type of information that is not very relevant for generation, such as grammatical explanations and feedback about their performance. All these requirement will be met by specifically designed tools that will be added to the programme.

3.1. Learners of a foreign language

Of these groups the first one is the most heterogeneous in two respects: the background knowledge they possessed and the kind of activities they perform. Usually they are classified according to the competence they have in the target language (learners of a foreign language for specific purposes are not included in this project). In this project this group has been divided into three sub-groups: elementary, intermediate and advanced, following the classification used by the language courses from which exercises have been taken. In principle it has been possible to adapt exercises from language courses of all three levels for English and Spanish; for German, whose computational grammar is more restricted at present, it was only possible to consider the adaptation of exercises of elementary and intermediate level.

This group of users should not be expected to have any substantial knowledge of linguistics, and this has two immediate consequences: first, the way grammatical explanations are presented must avoid hard terminology; second, only a very elementary type of information must be presented in the graphic structures of the sentences. This information should be restricted to basic notions such as *subect*, *direct object*, etc. Finally, this group of users differs from the others in the fact the some elements that make learning more attractive are often welcomed. One such element is the presentation of pictures and visual information in general, which will be experimented with the language for the weather. Another attractive element, the inclusion of a competitive factor such a score or time limits (cf. Higgins and Johns, 1984), will also be considered.

3.2. University students of linguistics

This group is expected to have a more solid knowledge of grammar and so it is unnecessary to avoid hard terminology, since this might actually be part of the contents of their university courses. What is crucial for this group of users is the analysis of linguistic forms and how this analysis is presented on the screen. KPML offers a tree-analysis of sentences, but other forms of graphical representation might also be required, such as boxes and underlining. It will also be necessary to include a menu where different types of information about the constituents can be switched on and off. For example, some users might want to work only with syntactic functions, while others might be interested in semantic functions like Actor or Affected. There can even be users who would like to see the three types of structure of the sentence according

to systemic-functional grammar: ideational, thematic and interpersonal. If more than one type of structure is to be displayed, it would be useful that they had different colours. Finally, morphological information should be presented in a more straight-forward form.

3.3. Spanish students of primary and secondary school

This group of users has similar needs to the group in the previous section, with the difference that their background knowledge is limited and so hard terminology should be avoided. Instead it will be necessary to present information using simple terms of traditional grammar, such as *subject*, *direct object*, etc. As in the previous group it will also be important to include more than one type of graphical representation of the structure (trees, boxes, underlining) as well as morphological information alongside syntactic functions. For this group only syntactic functions are needed, and all the information concerning semantic functions should be removed.

4. Creation of resources

A significant proportion of the time of the first year of the project has been devoted to the creation or improvement of linguistic resources. Linguistic resources refer here to the grammar and lexicon necessary for the generation of sentences.

In the case of Spanish the resources had to be created almost from scratch, since the only existing Spanish grammar for KPML turned out to be too limited (it was originally written to generate sentences belonging to a very specific topic: chemical substances). At the present moment the Spanish grammar has a degree of comprehensiveness similar to that of English (the English grammar contained in KPML is one the most complete computational grammars), although clause complexes are still to be refined. Since September 2002 there is a generation bank of Spanish with 138 sentences grouped in 8 regions of the grammar, each one exemplifying a different construction (see appendix B). 8 more regions providing about 150 further example sentences will be added very shortly.

The English grammar was complex enough to start with the design of some exercises right from the beginning of the project. But this process revealed that certain areas of the grammar required further refinement. The list of constructions that the initial grammar of English could not generate can be consulted in appendix C.

The German grammar was in a similar state to that of Spanish at the outset, but it was decided that any alterations of the grammar should be performed only by developers fully competent in that language. The type of activities that will be tested with German is therefore much more limited in range than those of English and Spanish, and probably will only include activities at an elementary level throughout the duration of the current project.

5. Description of the tools required for didactic purposes

After considering the potential users of KPML for language learning and the type of activities they are likely to demand, the following tools should, in my opinion, be added to the programme.

    a) An interface where the user can type in his/her answer to the exercises and the programme can examine it to provide feedback.

    b) An interface for the preparation of activities with sentences. It should be able to load sets of sentences and perform the followings operations:

-     hide some words

- show only the base form of certain words
- shuffle the words
- change sentence properties such as polarity, voice, tense and mood.

c) Different layers of information must be stored in the inquiries. The inquiries are the component in KPML responsible for choices in the grammar. These inquiries contain semantic explanations about the grammar, but in a technical style. It will be necessary to define new slots where simpler grammatical explanations can be stores, as well as other type of material such as visual information.

d) An interface for the preparation of activities with texts. This interface should be able to load stored texts and perform the operations mentioned in section 2.4.

e) A simpler interface for sentence generation and relations between clauses.

f) An interface for data searches. This interface should be able to load sets of examples and make searches based on the criteria mentioned in section 2.3. Finally, the results of the search should be displayed.

g) An interface that reads the output structure as KPML presents it at present and re-arranges it according to the users' preferences. The interface should include these options: type of graphical representation (tree, boxes, underlining), type of analysis (syntactic functions, morphological description, ideational structure, thematic structure, interpersonal structure).

6. Future work

After the first year of research the necessary linguistic resources for the creation of language activities have been created, possible groups of users have been established and their needs have been identified. This information has been used to sketch the main properties of future pedagogic tools that will be added to the generator.

The next six months will be devoted to the implementation of these tools and their testing. The initial design responds to a first evaluation of possible users' needs, but beta versions will be tested with students to make sure that the evaluation was correct or introduce the necessary changes after students' suggestions.

In the next six months I will also prepare activities for their testing with students. The activities will represent examples of what has been defined in this report as activities at or below sentence level and basic writing. The activities that involve large texts and grammatical analysis will probably be treated later on. But the collection of sentences for the creation of an English and a Spanish corpus will continue, and it is to be expected that after this six months they will contain enough data to proceed with their didactic exploitation.

REFERENCES

Biber-Conrad-Rippen (1998) *Corpus Linguistics. Investigating Language Structure in Use*. Cambridge: Cambridge University Press.

Granger, S. (1998) *Learner English on computer*. London, New York: Longman.

Higgins, J. And Johns, T. (1984) *Computers in language learning*. London: Collins ELT.

Hope-Taylor-Pusack (1984) *Using computers in teaching foreign languages*. Orlando: Harcourt Brace Jovanovich.

Johns, T. (1993) Date-driven Learning: An Update. *TELL & CALL 1993/2*, 2.

Kenning, M.J. and Kenning, M.M. (1984) *Introduction to computer assisted language teaching*. Oxford: Oxford University Press.

Levy, M. (1997) *Computer-assisted language learning: context and conceptualization*. Oxford: Clarendon Press.

Mann, W.C. and Thompson, S.A. Rethorical structure theory: a theory of text organization. In L. Polanyi, *The Structure of Discourse*. Norwood, NJ: Ablex Publishing Corporation.

McEnrey-Wilson-Baker (1997) Teaching and Language Corpora (TALC). *ReCall* 9.

Murphy, R. (2000a) *Essential Grammar in Use*. Cambridge: Cambridge University Press.

Murphy, R. (2000b) *English Grammar in Use*. Cambridge: Cambridge University Press.

Robinson, G. (1989) The CLCCS CALL Study: Methods, Error Feedback, Attitudes, and Achievment. In Smith, W. *Modern technology in foreign language education: applications and projects*. Lincolnwood, Ill: National Textbook.

Sánchez Lobato, J. (1999) *Español 2000. Nivel elemental*. Madrid: Sociedad General Española de Librería.

Schmied, J. (1999) Applying Contrastive Corpora in modern contrastive grammars: The Chemnitz Internet Grammar of English. In Hasselgard and Oksenfjell eds. *Out of Corpora. Studies in Honour for Stig Johansson.* Amsterdam: Rodopi, 21-30.

Stevens, V. (1986) Using LUCY/ELIZA as a means of facilitating communication in ESL, *TESOL Newsletter*, 20: 2, 13-14.

Underwood, J. (1984) *Linguistics, computers and the language teacher: a communicative approach*. Rowley, Mass: Newbury House Publishers.

Weizenbaum, J. (1976) *Computer Power and Human Reason*. Freeman.

Winograd, T. (1972) *Understanding Natural Language*. Academic Press (1984) Computer software for working with language. *Scientific American* 251: 130-45.

Witton, N. (1992) Answer Analysis, Feedback and On-Line Reference Help in CALL with Particular Reference to German. In Pennington & Stevens Computers in applied linguistics: An international perspective. Clevedon, Englad: Multilingual Matters.

APPENDIX A

Examples of exercises at or below the sentence level. The particular examples shown here have been adapted successfully to KPML. This means that exercises of this type can be created automatically using the KPML approach. The English exercises are *English Grammar in Use* (Murphy 2000), *Essential Grammar in Use* (Murphy 2000); the Spanish exercises are *Español 2000* (Sánchez Lobato y Fernández García 1999).

*Complete the sentences. Use a superlative (-**est** or **most** ...) + a preposition.*

1. It's a very nice room. It is *the nicest room in* the hotel
2. It's a very cheap restaurant. It's .... the town.
3. It was a very happy day. It was ....my life.
4. She's a very intelligent student. She .... the class.
5. It's a very valuable painting. It .... the gallery.
6. Spring is a very busy time for me. It .... the year.

Order the words

1. (the door/opened/I/quietly)   *I opened the door quietly*
2. (two letters/I/this morning /wrote)
3. (passed/Paul/easily/the exam)
4. (Ann/very well/French/doesn't speak)
5. (a lot of work/did/I/yesterday)
6. (London/do you know/well?)
7. (we/enjoyed/very much/the party)
8. (the problem/carefully/I/explained)

Conjugue los verbos que están entre paréntesis.

1. Nosotros (esperar) .... el autobús.
2. Usted (furmar) .... mucho.
3. Ellos (practicar) .... la lección.
4. Ella (contestar) .... la carta.
5. Vosotros (estudiar) .... la lección.
6. Yo (escuchar) .... la radio.
7. El profesor (explicar) .... la gramática.
8. Tú (preguntar) .... mucho.
9. Ustedes (hablar) .... muy bien español.
10. María (comprar) .... el periódico.

Ponga en plural las siguientes oraciones.

1. El avión es rápido. → Los aviones son rápidos.
2. La habitación está muy limpia.
3. Aquella chica es estudiante.
4. Este bolígrafo es azul.
5. La crisis económica es muy grave.
6. Este problema es muy difícil.

APPENDIX B

Generation bank of Spanish showing 138 sentences that illustrate constructions from different regions of the grammar. The computational input and the automatically generated structures of the sentences are available on the site: http://www.fb10.uni-bremen.de/anglistik/langpro/kpml/genbank/generation-bank.html.

Of the 15 examples in the listing that are reported as generating sentences not matching their targets, it is worth noting that almost all of these are caused by Spanish punctuation which was not supported adequately by the version of KPML used during the first year of the project. This deficit has already been corrected in the current version of KPML.

APPENDIX C

List of constructions that the present computational English grammar for KPML cannot generate. This presents a set of tasks that need to be addressed in the ongoing development of the English generation grammar and which were not presently identified.

- Combinations of two or more determiners. Ex.: ***My two*** *favourite actors./**All the other** options*.
- Comparative sentences in which the compared element is not noun group. Ex.: *It is easier for children than **for adults**./There more people than **I had expected**.*
- Wh-questions in which the wh-element is a constituent of a subordinated clause.Ex.: ***What do you think you are doing?***
- Comparative forms of adverbs and quantifiers modifying a noun. Ex.: *Run **faster**!/ There were **more** people.*
- Agreement between the subject and its attribute. Ex.: *My sister is **a nurse**./ My sisters are all **nurses**.*
- Combinations of modality and tense. Ex.: *Now I **have** to go/ Then I **had** to go.*
- Nouns that are always in the plural form. Ex.: ***These scissors*** *are not sharp*
- Combinations of two or more circumstances. Ex.: ***Yesterday*** *I worked **from nine to five**.*
- Time extents that are not realised by a prepositional phrase. Ex.: *He stayed there **all the morning**.*
- Particular negation of certain quantifiers. Ex.: ***Not many*** *people know this.*