SEMANTIC FUNCTIONS IN INSTRUCTIONAL TEXTS: A COMPARISON BETWEEN ENGLISH AND SPANISH

Susana Murcia Bielsa (susana.murcia@uam),
Universidad Autónoma de Madrid;
Michael O'Donnell (micko@wagsoft.com),
WagSoft Systems

1. Aims

This paper attempts to outline the range of 'semantic functions' which occur within instructional texts. By semantic functions, we mean either speech acts, or elements supporting these, such as purpose, condition, temporal, etc. Instructional texts are commonly used in everyday life and improving our understanding of these documents may lead to improvement in their production.

This exploration of semantic functions is intended as the first step of a larger program in the investigation of instructional texts. Subsequent steps will involve (i) the more detailed study of each of the semantic functions in turn, particularly in regard to the range of forms which realise each function, and (ii) explanation of the contextual factors which motivate a writer to choose between these forms.

Section 2 outlines what we mean by instructional texts, and Section 3 explains our notion of semantic function. Section 4 presents the data and methodology of the study, while section 5 provides the results. We summarise the study in section 6.

2. Instructional Texts

The object of our study are texts which define procedures. We focus on "Consumer Product Instructions" (CPIs), that is, instructional texts that accompany purchased household appliances. The length of these texts varies from one to ten or more pages. For the study we have placed the texts under three categories:

- *leaflet*, consisting of one single paper (either A4 or another size), possibly folded;
- *small booklet*, which contain less than 10 pages, and do not have a contents page;
- *large booklet*, with 10 or more pages and a contents page.

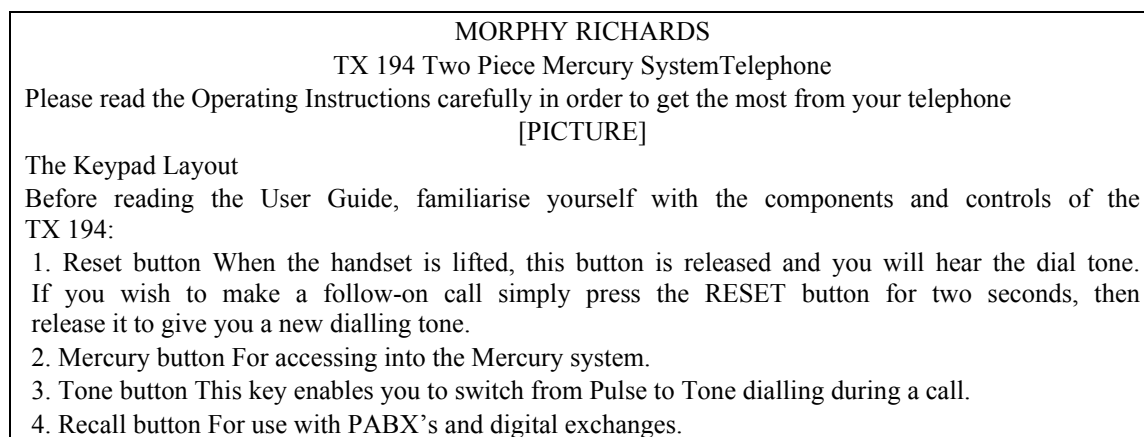Figure 1 shows a sample of an instructional text.

MORPHY RICHARDS

TX 194 Two Piece Mercury SystemTelephone

Please read the Operating Instructions carefully in order to get the most from your telephone

[PICTURE]

The Keypad Layout

Before reading the User Guide, familiarise yourself with the components and controls of the TX 194:

1. Reset button When the handset is lifted, this button is released and you will hear the dial tone. If you wish to make a follow-on call simply press the RESET button for two seconds, then release it to give you a new dialling tone.

2. Mercury button For accessing into the Mercury system.

3. Tone button This key enables you to switch from Pulse to Tone dialling during a call.

4. Recall button For use with PABX's and digital exchanges.

Figure 1: A Sample of an Instructional Text

3. Semantic Functions

We started with an interest in three functional elements of discourse, which have all previously been studied in relation to instructional texts[1]:

- *advice*: the writer tells the reader to do (or not do) something, also called directive in the studies referenced in the footnote:

    Press the RESET button

---

[1] See for instance Ciliberti (1990), Puglielli (1990), Delin et al. (1994, 1996), Paris & Scott (1994), Grote (1995), Carroll & Delin (1998), Murcia (1999, 2000), Murcia & Delin (2001).

- *informatives*: the writer provides the reader with information:
  > *Your telephone is guaranteed for one year*
- *purpose*: the writer tells the reader why some action was/is/will be necessary:
  > *To make a call, ...*

The first two of these are types of speech acts, while the third one is not typically considered a speech act, but rather is an element supporting a speech act. We thus introduce the term 'semantic function' to group together speech acts and elements which support them.

Our research program forced us to look at instructional texts as a whole. We thus needed to address the question: what other semantic functions constitute an instructional text? Apart from giving advice and information, and motivating these through purpose expressions, what does an instructional text need to do to be functional? The rest of this paper is our answer to this question.

4. Data and Methodology

Data: The first step in our study was data collection. We obtained a range of CPIs in both English and Spanish. This preliminary corpus included 12 texts, 6 texts in each language, varying in size, product, and manufacturer. All the texts were written in the original language, as translated texts may bear traces of the patterns of the original language. Each text was of approximately corresponding wordcount to one in the other language, to avoid biases due to variation in size. The final corpus consisted of 7,682 words for English, and 7,697 words for Spanish.

Analysis: Once the corpus was collected, the texts were scanned into the computer and analysed using the *Systemic Coder*[2], a semi-automatic software tool for coding text based on a coding scheme designed by the analyst, and which provides comparative and descriptive statistics from the corpus (see O'Donnell 1995).

We first used the tool's segmentation interface to divide our data into the units of study. The primary unit for our study is the finite clause, as this is the typical unit realising speech acts. However, the supporting elements of interest are usually realised by adjuncts, which may be clauses, but often are phrases. Our approach was thus to segment the data into independent clauses, and then separate off the adjuncts[3], which extend on (support) the semantics of these clauses.

The next step of the study then involved assigning semantic functions to each of these segments (using the Coder's tagging interface). We started with a very simple coding scheme, and as we encountered instances which were not covered, we added categories to the scheme. The model of the data was thus developed incrementally[4]. The resulting model of analysis is given below.

Once the data was coded, we used the statistical functions of the Coder to produce tables comparing different dimensions of the data, for instance, English vs. Spanish, and Document Size. The tables also indicate which of the differences between the data sets are statistically significant.

The coding model: The coding scheme, at the end of the study, was as follows:
- **Language**: *English* or *Spanish*;
- **Text Size**: *leaflet*, *small-booklet* or *large-booklet*
  The size of document containing the text segment.
- **Semantic Function**: *speech-act*, *supporting-unit, title, address, phatic-message*
  The general type of semantic function. The speech-act and supporting unit distinction was introduced above. We also found several types of text segments which did not fall into these categories, and were thus forced to add categories for the following:
  a) *Title*: document and section titles;
  b) *Address*: most documents sign off with the manufacturer's postal address;
  c) *Phatic-message*: the writer often includes a phatic message, such as *Dear Customer*, *Congratulations,* or *Gracias!*

In regard to speech-acts, we identified the following range of units:
- Informative (of the type *This is so*),
- Advice (*Do this!*), and

---

[2] Available free from: http://www.wagsoft.com/Coder/index.html.

[3] We include under adjuncts dependent clauses, as in "Do not open the lid <u>while the machine is running</u>".

[4] This corresponds to a Piagetian assimilation/ accommodation learning cycle: the model 'assimilates' as long as new instances are covered by the scheme, but we need to 'accomodate' (modify the scheme) when the instance is not covered.

- Warning (*Do not do this!*).
- Information seeking moves do not seem to play much of a role in CPIs, although there were 2 instances of rhetorical questions acting as conditions, e.g., *Does the fuse need replacing?*

In classifying the clausal adjuncts, we drew upon categories from Martin (1992). We identified the following supporting functions:
- Purpose (*To achieve this...*);
- Condition (*If this state exists...*);
- Temporal (*...after 3 minutes*);
- Location (*...towards the left*);
- Manner (*...in this way*);
- Instrument (*...using this instrument*);
- Reason (*...for this reason*);
- Explanation (*...as this will cause damage*).

The coding exposed some problem cases, such as how to differentiate (i) between Temporal and Condition, and (ii) among Purpose, Reason and Explanation.

Each of these functional units can be realised by a range of forms: for instance, Advice, while often realised by an imperative, can be realised by a declarative, e.g., *we recommend you do this*. See Murcia (2000) in regards to variation in advice expression and Murcia & Delin (2001) for purpose expressions. Future work will explore the forms of the other semantic functions, and motivate the choice between alternative forms.

5. Results

Segmentation resulted in 1,315 segments for English, and 1,011 for Spanish (the Spanish segments thus containing more words). These break down as shown in table 1. The '+' indicators on the left indicate how statistically significant the results are: '+' indicates significance at the 90% level; '++' indicates significance at the 95% level '+++' indicates significance at the 98% level.

	English (1315)	Spanish (1011)	
speech-act	45.4%	49.4%	++
supporting	45.5%	36.9%	+++
title	8.4%	12.6%	+++
phatic-unit	0.3%	0.6%	
address	0.4%	0.1%	

Table 1: Distribution of functional elements

General: Spanish contains a significantly higher percentage of speech-acts, and also titling. The explanation of this seems to be that English includes more supporting elements per speech-act. English uses on average one supporting clause per speech-act, while Spanish uses only 0.75 per speech-act.

	English	Spanish	
Purpose	13%	17%	+
Condition	19%	12%	+++
Location	25%	10%	+++
Temporal	16%	22%	+++
Manner	17%	17%	
Instrument	4%	9%	+++

Table 2: Support in English vs. Spanish

Supporting units: Exactly which supporting units are used in each language is shown in Table 2. The most significant differences between the two languages are the following:
- English uses the Location function more than twice as often as Spanish (25% vs. 10%);
- Although the use of Instrument is quite limited when compared to the other supporting functions, Spanish uses it more than twice as often as English (9% vs. 4%);
- Spanish uses the Temporal function significantly more frequently than English (22% vs. 16%);

3

- English uses Condition significantly more frequently than Spanish (19% vs. 12%);
- Spanish uses Purpose more frequently than English (17% vs. 13%), while Manner is used equally frequently in both languages.

Speech act types: In regards to the speech-acts, the division is as shown in table 3. The majority of the clauses in both languages perform the semantic function of Advice, as would be expected from any type of instruction. It seems that English CPIs include more Advice, while Spanish provides a higher degree of Informatives. While English uses slightly more Warnings, the difference is not significant.

	English (597)	**Spanish** (504)	
Advice	59%	48%	+++
Informative	32%	45%	+++
Warning	9%	7%	

Table 3: Speech-acts across languages.

Text Type: Unfortunately, we matched texts on word-count, rather than text-type. As it turned out, this did not mean we matched documents on type (leaflet vs. small-booklet vs. large-booklet). A comparison across languages was thus not possible. However, the results from within each language are interesting enough.

	Leaflet	**Large Booklet**	
speech-act	47%	56%	+++
supporting	53%	44%	+++

Table 4: Comparing English Text Sizes.

English large booklets use more speech-acts than supporting units (56% vs. 44%), while in the case of leaflets we find the opposite case, i.e., more supporting units than independent clauses (53% vs. 47%).

In Spanish, we found that leaflets use far more Warnings than small booklets (12% vs. 5% +++). Our guess is that explicit warnings are more to the point, and thus more common in a condensed text.

6. Summary and Conclusions

Conclusions: The study showed that consumer product instructions use a blend of advice, information-giving and warnings, Advice being the most common function in both English and Spanish. The analysis also showed that CPIs need a quite limited range of supporting functions, the most common of which are Location, Temporal, Manner, Condition, Purpose and Instrument.

When comparing the two languages, the analysis showed that English uses far more Advice than Spanish, and Spanish more Informatives. In addition, English includes more supporting adjuncts per speech-act, especially Location and Condition.

We also discovered that the longer the document length, the higher the degree of informatives in the booklet. Shorter texts can consist of mostly directives/advice, while longer texts offer additional background information.

Applications: The results of this study have several possible applications. Firstly, human translators of CPIs can be made aware of the way in which these texts differ in each language, and use the study to guide their translation. Secondly, machine production of CPIs, either by machine translation or text generation, can be informed by the patterns reported here.

Future Work: This study was organised as the first step in a longer project. Subsequent studies will explore the nature of each semantic function, for instance, (i) what range of expressions are available for each function, and (ii) in what context is each form preferred. We also need to increase our corpus size on this study to verify these preliminary results.

BIBLIOGRAPHY/REFERENCES

Carroll, T. & Delin, J.: "Written Instructions in Japanese and English: a Comparative Analysis". *Pragmatics*, 8 (3); pp.339-385, 1998.

Ciliberti, A.: "Instructions for Use: A Macrotextual and Stylistic Analysis". In: Halliday, M.A.K., J. Gibbons and H. Nicholas (Eds.) *Learning, Keeping and Using Language. Selected Papers from the 8th World Congress of Applied Linguistics*. Sydney, 16-21 August 1987, (vol. II). Amsterdam: John Benjamins, 1990; pp. 299-313.

Delin, J., Hartley, A., Paris, C., Scott, D. & Vander Linden, K. (1994) "Expressing Procedural Relationships in Multilingual Instructions". *Proceedings of 7th International Workshop on Natural Language Generation*. Kennebunkport, Maine, June 1994; pp. 61-70.

Delin, J., Hartley, A. & Scott, D.: "Towards a Contrastive Pragmatics: Syntactic Choice in English and French Instructions". *Language Sciences*, 18 (3-4), 1996; pp. 897-931.

Grote, B.: "Expressing procedural relationships in German instructional texts". Technical Report, Information Technology Research Institute, University of Brighton, ITRI-95-6, Brighton, 1995.

Martin, J.R.: *English Text: System and Structure*. Amsterdam: John Benjamins, 1992.

Murcia Bielsa, S.: *Instructional Texts in English and Spanish: a Contrastive Study*. Unpublished Ph.D. thesis, Universidad de Córdoba, 1999.

Murcia Bielsa, S.: "The choice of directive expressions in English and Spanish instructions: a semantic network". In: Eija Ventola (Ed.) *Discourse And Community. Doing Functional Linguistics*. Language in Performance 21. Tubingen: Gunter Narr Verlag, 2000.

Murcia Bielsa, S. & Delin, J.: "Expressing the notion of purpose in English and Spanish instructions". *Functions of Language* 8(1), 2001; pp. 33-56.

O'Donnell, M.: "From Corpus to Codings: Semi-Automating the Acquisition of Linguistic Features". *Proceedings of the AAAI Spring Symposium on Empirical Methods in Discourse Interpretation and Generation, March 27-29, 1995*. Stanford University, California.

Paris, C. & Scott, D.: "Stylistic Variation in Multilingual Instructions". *Proceedings of the 7th International Workshop on Natural Language Generation*, Kennebunkport, Maine, June 1994; pp. 45-52.

Puglielli, A.: "Instructions for Use: From Macro to Microlinguistic Analysis". In: M.A.K. Halliday, J. Gibbons and H. Nicholas (Eds.) *Learning, Keeping and Using Language. Selected Papers from the 8th World Congress of Applied Linguistics*. Sydney, 16-21 August 1987, (vol. II). Amsterdam: John Benjamins, 1990; pp. 315-332.