SPLAT: A sentence-plan authoring tool

Bruce Jakeway and Chrysanne DiMarco[t]

[t]Department of Computer Science
University of Waterloo
Waterloo, Ontario N2L 3G1, Canada
cdimarco@logos.uwaterloo.ca
Voice: +1 519 888 4443
Fax: +1 519 885 1208

ABSTRACT

SPLAT (Sentence Plan Language Authoring Tool) is an authoring tool intended to facilitate the creation of sentence-plan specifications for the Penman natural language generation system. SPLAT uses an example-based approach in the form of *sentence-plan templates* to aid the user in creating and maintaining sentence plans. SPLAT also contains a *sentence bank*, a user-extensible collection of sentence plans annotated in various ways. The sentence bank can be searched for candidate plans that can then be used in the creation of new sentence plans specific to the domain of interest. SPLAT's graphical environment provides additional support to the user in the form of menu-driven access to Penman's linguistic resources and management of partially built sentence plans.

1 Introduction

As natural language generation systems become more complex and sophisticated, the mode of input to these systems is becoming correspondingly more difficult to specify and manage. Currently, sentence plans must be created by experts who are very knowledgeable about both linguistic theory and the characteristics of the particular generation system. In order to facilitate the growth of natural language generation research, systems should be able to handle more complex input, but at the same time should be more accessible to non-experts. To accomplish these goals requires a facility that aids the user in creating and maintaining the desired input specifications in a principled and convenient way. Such a facility should have both theoretical knowledge of the allowable form and content of the input specifications, and the practical ability to ensure that only syntactically correct specifications are entered. Ideally, it should also maintain a library of sample input specifications so that the user need not reconstruct existing specifications, but can construct new ones from an existing set. As well, such a facility should provide access to a range of resources, such as grammars and other linguistic resources that can aid the user in input development. SPLAT is a first attempt to provide such a facility for the Penman generation system in the form of an authoring tool for Sentence Plan Language (SPL).

2 The Penman system: The need for authoring

2.1 The input to Penman: Sentence Plan Language

The Penman generation system [Penman, 1989] is one of the most comprehensive natural language generation systems in the world. It contains a very large systemic functional grammar of English, the *Nigel grammar*, and an extensive semantic ontology, the *upper model*. Input to the Penman system is defined by the Sentence Plan Language [Penman, 1991] and given in the form of SPL plans. The system processes the SPL input by querying different knowledge resources, including the Nigel grammar, the upper model, and a domain model, eventually producing a realization of the SPL plan in the form of an English sentence.

Each SPL plan contains one or more head concepts from the upper model and a variable that enables the plan to be referred to by other SPL plans. The head concept can

be modified by a number of keywords, signalling the underlying grammar to generate different sentence patterns. For example, one common type of keyword is a *relation*, whose value may be another SPL plan, or a reference to such a plan.

2.2 Why using SPL is hard

One of the major difficulties in learning to use Penman has been acquiring the expertise to construct SPL plans. The vast number of possible keywords and values, coupled with the absence of any facility for storing and searching these items, made the task of building SPL plans very frustrating for the novice Penman user. In contrast, the experienced SPL developer would often draw on knowledge of previously constructed SPL plans in order to recycle bits of partial SPL plans, but had no convenient way to store and access this information. Various resources, the upper model, for instance, play an essential role in providing knowledge to guide the development of the SPL plan, but were virtually inaccessible to all but the Penman expert. Support for managing the construction of SPL plans and accessing the necessary resources in a systematic and user-friendly manner was almost completely lacking.

3 An authoring tool for SPL

SPLAT is a Sentence Plan Language Authoring Tool that has been developed as part of the HealthDoc project [Jakeway, 1995, Di-Marco *et al.*, 1995] and aims to address many of the earlier difficulties in building the input SPL specifications for Penman. SPLAT allows the user to create SPL plans in a supportive on-line environment: a graphical, menu-driven interface provides guidance on the allowable structure of SPL plans and access to the various Penman resources, such as the upper model and the generator itself. As well, SPLAT provides an extensible bank of representative sentences and their SPL structures, from which the user can create new sentence plans. An important feature of SPLAT is that it provides a new view of the upper model.

3.1 The modelling approach

SPLAT draws on some aspects of modelling theory in helping the user build SPL plans, specifically, by giving examples of SPL-plan *templates* and prefabricated SPL plans. The templates supply most of the format specific to a particular kind of SPL plan, thus reducing the need for the user to memorize SPL-plan syntax. This feature also helps reduce the possibility of syntactic errors in plan building. In addition, having a template display the allowable components of a particular type of SPL plan guides the user in exploring the Penman system and in learning how the different parts of an SPL plan interact.

Previously constructed SPL plans provide models that may be modified or incorporated into the new plan. This extensible example set, the *sentence bank*, provides positive examples of how to construct SPL-plan templates. The user can retrieve the plan for a particular token in the sentence bank and modify it, or use it to aid in the construction of a new plan.

3.2 Remodelling the upper model

The Penman upper model, a classification of various semantic concepts, has traditionally been divided into three disjoint concept hierarchies, a high-level split between the major semantic abstractions of English: *processes, objects,* and *qualities*. Processes can be thought of as verbs and other relational concepts, objects as nouns, and qualities as modifiers of processes and objects, *i.e.*, adverbs and adjectives. However, early in the construction of SPLAT, it was noted that processes should be divided into two categories, those which modify the ideational content of a sentence and those which dictate the textual structure of a sentence. The former category describes most of the process hierarchy (*i.e.*, verbs and most relations), whereas the latter describes the logical and rhetorical relations. Adding this category to SPLAT takes the upper model closer to the semantic classification described by [Matthiessen, 1991], with textual functions being separated from ideational functions.

6

3.3 Building SPL templates

SPLAT provides template forms for each type of SPL plan: relations, processes, objects, and qualities. The user need only fill in appropriate values on the selected template. As well, for processes, the template is not a static structure, but changes for each type of process according to its roles, which are retrieved from the underlying knowledge base. For example, the template for a verbal process will display the roles relevant to this type of process: *sayer*, *addressee*, and *saying*. For each kind of template, SPLAT provides most of the roles necessary to construct this type of SPL plan.

Each template also provides a facility to add keywords and values that are not present on the form. The template also gives the user access to a number of different tools and resources, including the actual SPL that gets constructed, and the generator's output from the constructed SPL. At any point in the development process, the user can choose to produce the partially built SPL plan or generate (through Penman) the English realization of the current structure.

3.4 The sentence bank

SPLAT stores the pre-built SPL plans for a set of sample sentences in a *sentence bank*. Each token of a sentence in the sentence bank is connected back to the SPL-plan template associated with it; *i.e.*, the templates for a particular sentence and for its components are directly accessible from the sentence bank. Users can search the sentence bank to find examples of a particular sentence or partial sentence pattern. The corresponding SPL-plan template can then be used as the model or component of the new SPL plan being built. As users develop their own SPL plans, they can add them to the sentence bank by choosing the *annotation* feature on the current template.

3.4.1 The purpose of annotations

Each word in the sentence bank is annotated with up to five levels of annotation: spelling, lexical item, part of speech, grammatical function, and upper-model con-

cept. The spelling corresponds to the actual spelling of the word in the sentence and is retrieved from the generator output. The lexical item is the lexical unit used by Penman to generate the word, derived from the SPL input. The part-of-speech annotation for a particular word is derived from Brill's [1994] part-of-speech tagger. Before a sentence is entered into the sentence bank, it is passed to the tagger to determine the part of speech of each word in the sentence. The grammatical function of a word is derived from the output of the Penman generator.[1] The upper-model concept attributed to the word is retrieved from the SPL input. Not all words of a sentence will be annotated at each level, as some annotation levels might not apply to a particular word.

Table 1 shows how SPLAT would annotate a sample sentence. Notice that the semantically more important words have full annotations, whereas the support words do not have lexical or conceptual information. These support words are grouped with their associated semantically meaningful words into tokens. These tokens correspond to SPL plans. For example, the word *were* is an auxiliary to the word produced from the concept *CREATIVE-MATERIAL-ACTION* because of tense requirements. (The cryptic syntactic part-of-speech tags are from the Penn Treebank tagset.) In the sentence bank, each token is presented as a unit that is linked to the underlying SPL-plan template, so that the template can be edited.

3.4.2 Searching the sentence bank

The sentence bank contains a number of sample sentences which illustrate various types of SPL plans and constructions. The user may want to make use of existing plans, either using them to guide the construction of new plans, or modifying them to create new ones. To determine which sentence plan to use, the user searches the sentence bank with a pattern indicating the desired settings

---

[1] SPLAT uses KPML [Bateman, 1995] version 0.8 or later to retrieve these values, as the KPML generator can easily return the full structure output of a sentence, whereas this information is much harder to retrieve from the standard Penman generator.

spelling	lexical item	part-of-speech	grammatical	semantic
Those	·	DT	DEICTIC	
people	PERSON	NNS	THING	PERSON
were		VBD	TEMPO1	
making	MAKE	VBG	VOICE	CREATIVE-MATERIAL-ACTION
a		DT	DEICTIC	
new	NEW	JJ	QUALITY	NEW
ship	SHIP	NN	THING	SHIP

Table 1: **Sentence bank annotations.** Example of annotations for the sentences *Those people were making a new ship.*

for the annotation levels. SPLAT will retrieve all the sentences in the sentence bank which match the pattern. For example, if the sentence in Table 1 was in the sentence bank, and if the pattern specified that a word with a syntactic part-of-speech *DT* (determiner) was to be followed by a word whose lexical item was *PERSON*, then it would be retrieved. If, however, the pattern was the word *those*, followed by any number of words, followed by a word with a concept matching *BELIEVE*, then the sentence would not be retrieved.

4 Conclusion

SPLAT provides a graphical environment where users can enter the SPL-plan specifications for the Penman generation system. This environment, including its SPL-plan templates and sentence bank, provides an easy way for Penman users, even novices, to create and maintain SPL plans. The SPL-plan templates display the necessary keywords for the particular type of plan, and SPLAT automatically enforces correct SPL syntax. The sentence bank provides the user with a convenient method for saving SPL plans and indexing them for later reuse. Currently, a graphical upper-model browsing tool has been implemented and other resources are being developed.

References

[Bateman, 1995] John A. Bateman. KPML: The KOMET-Penman multilingual resource development environment. In *Proceedings of the Fifth European Workshop on Natural Language Generation*, pages 219–222, Leiden, May 1995. Social and Behavioural Sciences, University of Leiden.

[Brill, 1994] Eric Brill. Some advances in transformation-based part of speech tagging. In *Proceedings of the 12th National Conference on Artificial Intelligence*, pages 722–727, Seattle, 1994.

[DiMarco et al., 1995] Chrysanne DiMarco, Graeme Hirst, Leo Wanner, and John Wilkinson. HealthDoc: Customizing patient information and health education by medical condition and personal characteristics. In *Workshop on Artificial Intelligence in Patient Education*, pages 59–71, Glasgow, August 1995.

[Jakeway, 1995] Philip Bruce Jakeway. SPLAT: A sentence plan authoring tool for natural language generation. Master's thesis, University of Waterloo, 1995.

[Matthiessen 1991] Christian Matthiessen. Lexico(grammatical) choice in text generation. In Cécile Paris; William R. Swartout; and William C. Mann, editors, *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, pages 249–292. Kluwer Academic Publishers, Norwell, Massachusetts, 1991.

[Penman, 1989] Penman Natural Language Generation Group. *The Penman Primer, The Penman User Guide*, and *The Penman Reference Manual*. Information Sciences Institute, 1989.

[Penman, 1991] Penman Natural Language Group. *The Penman SPL Guide*. Information Sciences Institute. Unpublished manuscript, 1991.