Statistical Acquisition of Content Selection Rules for Natural Language Generation

Pablo A. Duboue and **Kathleen R. McKeown**

Department of Computer Science
Columbia University
{pablo,kathy}@cs.columbia.edu

Abstract

A Natural Language Generation system produces text using as input semantic data. One of its very first tasks is to decide which pieces of information to convey in the output. This task, called Content Selection, is quite domain dependent, requiring considerable re-engineering to transport the system from one scenario to another. In this paper, we present a method to acquire content selection rules automatically from a corpus of text and associated semantics. Our proposed technique was evaluated by comparing its output with information selected by human authors in unseen texts, where we were able to filter half the input data set without loss of recall.

1 Introduction

CONTENT SELECTION is the task of choosing the right information to communicate in the output of a Natural Language Generation (NLG) system, given semantic input and a communicative goal. In general, Content Selection is a highly domain dependent task; new rules must be developed for each new domain, and typically this is done manually. Morevoer, it has been argued (Sripada et al., 2001) that Content Selection is the most important task from a user's standpoint (i.e., users may tolerate errors in wording, as long as the information being sought is present in the text).

Designing content selection rules manually is a tedious task. A realistic knowledge base contains a large amount of information that could potentially be included in a text and a designer must examine a sizable number of texts, produced in different situations, to determine the specific constraints for the selection of each piece of information.

Our goal is to develop a system that can automatically acquire constraints for the content selection task. Our algorithm uses the information we learned from a corpus of desired outputs for the system (i.e., human-produced text) aligned against related semantic data (i.e., the type of data the system will use as input). It produces constraints on every piece of the input where constraints dictate if it should appear in the output at all and if so, under what conditions. This process provides a filter on the information to be included in a text, identifying all information that is potentially relevant (previously termed *global focus* (McKeown, 1985) or *viewpoints* (Acker and Porter, 1994)). The resulting information can be later either further filtered, ordered and augmented by later stages in the generation pipeline (e.g., see the spreading activation algorithm used in ILEX (Cox et al., 1999)).

We focus on descriptive texts which realize a single, purely informative, communicative goal, as opposed to cases where more knowledge about speaker intentions are needed. In particular, we present experiments on biographical descriptions, where the planned system will generate short paragraph length texts summarizing important facts about famous people. The kind of text that we aim to generate is shown in Figure 1. The rules that we aim to acquire will specify the kind of information that is typically included in any biography. In some cases, whether

Actor, born Thomas Connery on August 25, 1930, in Fountainbridge, Edinburgh, Scotland, the son of a truck driver and charwoman. He has a brother, Neil, born in 1938. Connery dropped out of school at age fifteen to join the British Navy. Connery is best known for his portrayal of the suave, sophisticated British spy, James Bond, in the 1960s. ...

Figure 1: Sample Target Biography.

the information is included or not may be conditioned on the particular values of known facts (e.g., the *occupation* of the person being described —we may need different content selection rules for artists than politicians). To proceed with the experiments described here, we acquired a set of semantic information and related biographies from the Internet and used this corpus to learn Content Selection rules.

Our main contribution is to analyze how variations in the data influence changes in the text. We perform such analysis by splitting the semantic input into clusters and then comparing the language models of the associated clusters induced in the text side (given the alignment between semantics and text in the corpus). By doing so, we gain insights on the relative importance of the different pieces of data and, thus, find out which data to include in the generated text.

The rest of this paper is divided as follows: in the next section, we present the biographical domain we are working with, together with the corpus we have gathered to perform the described experiments. Section 3 describes our algorithm in detail. The experiments we perform to validate it, together with their results, are discussed in Section 4. Section 5 summarizes related work in the field. Our final remarks, together with proposed future work conclude the paper.

2 Domain: Biographical Descriptions

The research described here is done for the automatic construction of the Content Selection module of PROGENIE (Duboue and McKeown, 2003a), a biography generator under construction. Biography generation is an exciting field that has attracted practitioners of NLG in the past (Kim et al., 2002; Schiffman et al., 2001; Radev and McKeown, 1997; Teich and Bateman, 1994). It has the advantages of being a constrained domain amenable to current generation approaches, while at the same time offering more possibilities than many constrained domains, given the variety of styles that biographies exhibit, as well as the possibility for ultimately generating relatively long biographies.

We have gathered a resource of text and associated knowledge in the biography domain. More specifically, our resource is a collection of human-produced texts together with the knowledge base a generation system might use as input for generation. The knowledge base contains many pieces of information related to the person the biography talks about (and that the system will use to generate that type of biography), not all of which necessarily will appear in the biography. That is, the associated knowledge base is not the semantics of the *target text* but the larger set[1] of all things that could possibly be said about the person in question. The intersection between the input knowledge base and the semantics of the target text is what we are interested in capturing by means of our statistical techniques.

To collect the semantic input, we crawled 1,100 HTML pages containing celebrity fact-sheets from the *E! Online* website.[2] The pages comprised information in 14 categories for actors, directors, producers, screenwriters, etc. We then proceeded to transform the information in the pages to a frame-based knowledge representation. The final corpus contains 50K frames, with 106K frame-attribute-value triples, for the 1,100 people mentioned in each fact-sheet. An example set of frames is shown in Figure 3.

The text part was mined from two different websites, `biography.com`, containing typical biographies, with an average of 450 words each; and `imdb.com`, the Internet movie database, 250-word average length biographies. In each case, we obtained the semantic input from one website and a separate biography from a second website. We linked the two resources using techniques from *record linkage* in census statistical analysis (Fellegi and Sunter, 1969). We based our record linkage on the *Last Name, First Name,* and *Year of Birth* attributes.

[1]The semantics of the text normally contain information not present in our semantic input, although for the sake of Content Selection is better to consider it as a "smaller" set.

[2]`http://www.eonline.com`

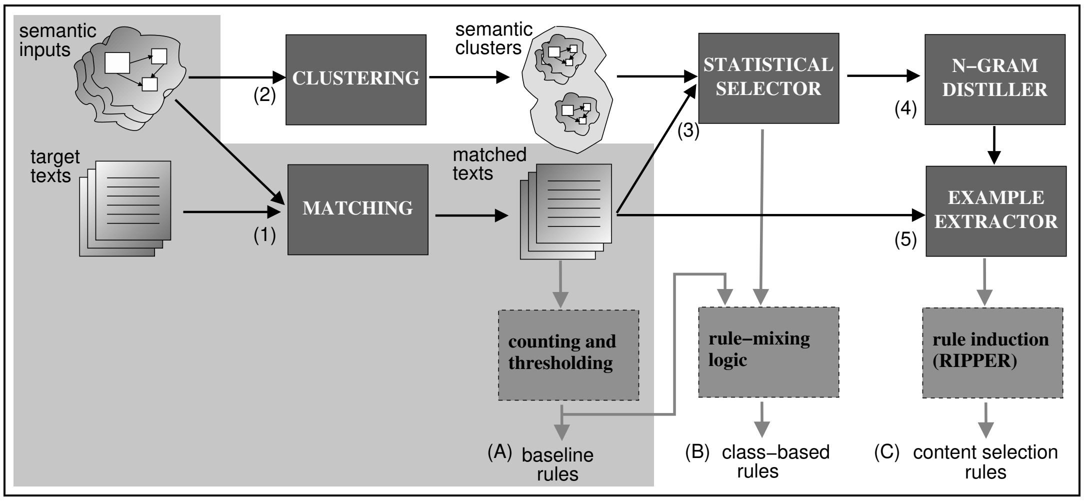

Figure 2: Our proposed algorithm, see Section 3 for details.

3 Methods

Figure 2 illustrates our two-step approach. In the first step (shaded region of the figure), we try to identify and solve the easy cases for Content Selection. The easy cases in our task are pieces of data that are copied verbatim from the input to the output. In biography generation, this includes names, dates of birth and the like. The details of this process are discussed in Section 3.1. After these cases have been addressed, the remaining semantic data is clustered and the text corresponding to each cluster post-processed to measure degrees of influence for different semantic units, presented in Section 3.2. Further techniques to improve the precision of the algorithm are discussed in Section 3.3.

Central to our approach is the notion of *data paths* in the semantic network (an example is shown in Figure 3). Given a frame-based representation of knowledge, we need to identify particular pieces of knowledge inside the graph. We do so by selecting a particular frame as the *root* of the graph (the person whose biography we are generating, in our case, doubly circled in the figure) and considering the paths in the graph as identifiers for the different pieces of data. We call these *data paths*. Each path will identify a *class* of values, given the fact that some attributes are list-valued (e.g., the **relative** attribute in the figure). We use the notation $\langle \texttt{attribute}_1\, \texttt{attribute}_2 \ldots \texttt{attribute}_n \rangle$ to denote data paths.

3.1 Exact Matching

In the first stage (cf. Fig. 2(1)), the objective is to identify pieces from the input that are copied verbatim to the output. These types of verbatim-copied anchors are easy to identify and they allow us do two things before further analyzing the input data: remove this data from the input as it has already been selected for inclusion in the text and mark this piece of text as a part of the input, not as actual text.

The rest of the semantic input is either verbalized (e.g., by means of a verbalization rule of the form $\langle \texttt{brother age} \rangle < 35 \Rightarrow$ *"young"*) or not included at all. This situation is much more challenging and requires the use of our proposed statistical selection technique.

3.2 Statistical Selection

For each class in the semantic input that was not ruled out in the previous step (e.g., $\langle \texttt{brother age} \rangle$), we proceed to cluster (cf. Fig. 2(2)) the possible values in the path, over all people (e.g., $[1 \leq age \leq 24]; [25 \leq age \leq 50]; [51 \leq age \leq 90]$ for age). Clustering details can be found in (Duboue and McKeown, 2003b). In the case of free-text fields, the top level, most informative terms,[3] are picked and used for the clustering. For example, for *"played an insecure young resident"* it would be [*played*, *insecure*, *resident*].

Having done so, the texts associated with each

[3]We use the maximum value of the TF*IDF weights for each term in the whole text collection. That has the immediate effect of disregarding stop words.

cluster are used to derive language models (in our case we used bi-grams, so we count the bi-grams appearing in all the biographies for a given cluster —e.g., all the people with age between 25 and 50 years old, $[25 \leq age \leq 50]$).

We then measure the variations on the language models produced by the variation (clustering) on the data. What we want is to find a change in word choice correlated with a change in data. If there is no correlation, then the piece of data which changed should not be selected by Content Selection.

In order to compare language models, we turned to techniques from adaptive NLP (i.e., on the basis of genre and type distinctions) (Illouz, 2000). In particular, we employed the *cross entropy*[4] between two language models M_1 and M_2, defined as follows (where $P_M(m)$ is the probability that M assigns to the n-gram m):

$$CE(M_1, M_2) = -\sum_i P_{M_1}(i) \log P_{M_2}(i) \quad (1)$$

Smaller values of $CE(M_1, M_2)$ indicate that M_1 is more similar to M_2. On the other hand, if we take M_1 to be a model of randomly selected documents and M_2 a model of a subset of texts that are associated with the cluster, then a greater-than-chance CE value would be an indicator that the cluster in the semantic side is being correlated with changes in the text side.

We then need to perform a sampling process, in which we want to obtain CE values that would represent the null hypothesis in the domain. We sample two arbitrary subsets of k elements each from the total set of documents and compute the CE of their derived language models (these CE values constitute our control set). We then compare, again, a random sample of size k from the cluster against a random sample of size k from the difference between the whole collection and the cluster (these CE values constitute our experiment set). To see whether the values in the experiment set are larger (in a stochastic fashion) than the values in the control set, we employed the Mann-Whitney U test (Siegel and Castellan Jr., 1988) (cf. Fig. 2(3)). We performed 20 rounds of sampling (with $k = 5$) and tested at the 0.05 significance level. Finally, if the cross-entropy values for the experiment set are larger than for the control set, we can infer that the values for that semantic cluster do influence the text. Thus, a positive U test for any data path was considered as an indicator that the data path should be selected.

[4]Other metrics would have been possible, in as much as they measure the similarity between the two models.

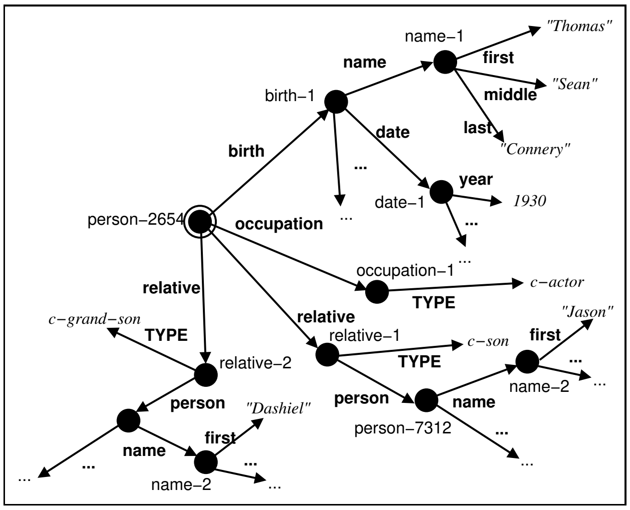

Figure 3: A frame-based knowledge representation, containing the triples (`person-2654`, **occupation**, `occupation-1`), (`occupation-1`, **TYPE**, *c-actor*), ... Note the list-valued attribute **relative**.

Using simple thresholds and the U test, *class-based* content selection rules can be obtained. These rules will select or unselect each and every instance of a given data path at the same time (e.g., if ⟨`relative person name first`⟩ is selected, then both *"Dashiel"* and *"Jason"* will be selected in Figure 3). By counting the number of times a data path in the exact matching appears in the texts (above some fixed threshold) we can obtain **baseline** content selection rules (cf. Fig. 2(A)). Adding our statistically selected (by means of the cross-entropy sampling and the U test) data paths to that set we obtain **class-based** content selection rules (cf. Fig. 2(B)). By means of its simple algorithm, we expect these rules to overtly over-generate, but to achieve excellent coverage. These class-based rules are relevant to the KR concept of *Viewpoints* (Acker and Porter, 1994);[5] we extract a slice of the knowledge base that

[5]they define them as *a coherent sub-graph of the knowledge base describing the structure and function of objects, the change made to objects by processes, and the temporal attributes and temporal decompositions of processes*.

is relevant to the domain task at hand.

However, the expressivity of the class-based approach is plainly not enough to capture the idiosyncrasies of content selection: for example, it may be the case that children's names may be worth mentioning, while grand-children's names are not. That is, in Figure 3, ⟨`relative person name first`⟩ is dependent on ⟨`relative TYPE`⟩ and therefore, all the information in the current instance should be taken into account to decide whether a particular data path and it values should be included or not. Our approach so far simply determines that an attribute should always be included in a biography text. These examples illustrate that content selection rules should capture cases where an attribute should be included only under certain conditions; that is, only when other semantic attributes take on specific values.

3.3 Improving Precision

We turned to `ripper`[6] (Cohen, 1996), a supervised rule learner categorization tool, to elucidate these types of relationships. We use as features a flattened version of the input frames,[7] plus the actual value of the data in question. To obtain the right label for the training instance we do the following: for the exact-matched data paths, matched pieces of data will correspond to positive training classes, while unmatched pieces, negative ones. That is to say, if we know that (⟨`brother age`⟩ ,26) and that 26 appears in the text, we can conclude that the data of this particular person can be used as a positive training instance for the case (⟨`age`⟩ ,26). Similarly, if there is no match, the opposite is inferred.

For the U-test selected paths, the situation is more complex, as we only have clues about the importance of the data path as a whole. That is, while we know that a particular data path is relevant to our task (biography construction), we don't know with which values that particular data path is being verbalized. We need to obtain more information from the sampling process to be able to identify cases in which we believe that the relevant data path has been verbalized.

To obtain finer grained information, we turned to a *n-gram distillation* process (cf. Fig. 2(4)), where the most significant n-grams (bi-grams in our case) were picked during the sampling process, by looking at their overall contribution to the CE term in Equation 1. For example, our system found the bi-grams $\boxed{\textit{screenwriter director}}$ and $\boxed{\textit{has screenwriter}}$[8] as relevant for the cluster (⟨`occupation TYPE`⟩, `c-writer`), while the cluster (⟨`occupation TYPE`⟩, {`c-comedian`, `c-actor`}) will not include those, but will include $\boxed{\textit{sitcom Time}}$ and $\boxed{\textit{Comedy Musical}}$. These n-grams thus indicate that the data path (⟨`occupation TYPE`⟩, is included in the text; a change in value does affect the output. We later use the matching of these n-grams as an indicator of that particular instance as being selected in that document.

Finally, the training data for each data path is generated. (cf. Fig. 2(5)). The selected or unselected label will thus be chosen either via direct extraction from the exact match or by means of identification of distiled, relevant n-grams. After `ripper` is run, the obtained rules are our sought **content selection rules** (cf. Fig. 2(5)).

4 Experiments

We used the following experimental setting: 102 frames were separated from the full set together with their associated 102 biographies from the `biography.com` site. This constituted our development corpus. We further split that corpus into development training (91 people) and development test and hand-tagged the 11 document-data pairs.

The annotation was done by one of the authors, by reading the biographies and checking which triples (in the RDF sense, (frame, slot, value)) were actually mentioned in the text (going back and forth to the biography as needed). Two cases required special attention. The first one was *aggregated information*, e.g., the text may say *"he received three*

[6] We chose `ripper` to use its set-valued attributes, a desirable feature for our problem setting.

[7] The flattening process generated a large number of features, e.g., if a person had a grandmother, then there will be a "grandmother" column for every person. This gets more complicated when list-valued values are taken into play. In our biographies case, an average-sized 100-triples biography spanned over 2,300 entries in the feature vector.

[8] Our bi-grams are computed after stop-words and punctuation is removed, therefore these examples can appear in texts like *"he is an **screenwriter**,**director**,... "* or *"she **has** an **screenwriter** award..."*

Grammys" while in the semantic input each award was itemized, together with the year it was received, the reason and the type (Best Song of the Year, etc.). In that case, only the **name** of award was selected, for each of the three awards. The second case was factual errors. For example, the biography may say the person was born in MA and raised in WA, but the fact-sheet may say he was born in WA. In those cases, the *intention* of the human writer was given priority and the place of birth was marked as selected, even though one of the two sources were wrong. The annotated data total 1,129 triples. From them, only 293 triples (or a 26%) were verbalized in the associated text and thus, considered selected. That implies that the "select all" tactic ("select all" is the only trivial content selection tactic, "select none" is of no practical value) will achieve an F-measure of 0.41 (26% prec. at 100% rec.).

Following the methods outlined in Section 3, we utilized the training part of the development corpus to mine Content Selection rules. We then used the development test to run different trials and fit the different parameters for the algorithm. Namely, we determined that filtering the bottom 1,000 TF*IDF weighted words from the text before building the n-gram model was important for the task (we compared against other filtering schemes and the use of lists of stop-words). The best parameters found and the fitting methodology are described in (Duboue and McKeown, 2003b).

We then evaluated on the rest of the semantic input (998 people) aligned with one other textual corpus (`imdb.com`). As the average length-per-biography are different in each of the corpora we worked with (450 and 250, respectively), the content selection rules to be learned in each case were different (and thus, ensure us an interesting evaluation of the learning capabilities). In each case, we split the data into training and test sets, and hand-tagged the test set, following the same guidelines explained for the development corpus. The linkage step also required some work to be done. We were able to link 205 people in `imdb.com` and separated 14 of them as the test set.

The results are shown in Table 1[9]. Several things can be noted in the table. The first is that `imdb.com` represents a harder set than our development set. That is to expect, as `biography.com`'s biographies have a stable editorial line, while `imdb.com` biographies are submitted by Internet users. However, our methods offer comparable results on both sets. Nonetheless, the tables portray a clear result: the **class-based** rules are the ones that produce the best overall results. They have the highest F-measure of all approaches and they have high recall. In general, we want an approach that favors recall over precision in order to avoid losing any information that is necessary to include in the output. Overgeneration (low precision) can be corrected by later processes that further filter the data. Further processing over the output will need other types of information to finish the Content Selection process. The **class-based** rules filter-out about 50% of the available data, while maintaining the relevant data in the output set.

[9]We disturbed the dataset to obtain some cross-validation over these figures, obtaining a std dev. of 0.02 for the F*, the full details are given in (Duboue and McKeown, 2003b).

```
SELECT ⟨award subtitle⟩
   IF ⟨occupation₁ TYPE⟩ = director AND
      ⟨education₂ place country⟩ = USA AND
      ⟨award₅ title⟩ ⊅ Festival
```

Figure 4: Example rule, from the `ripper` output. It says that the subtitle of the award (e.g., "Best Director", for an award with title "Oscar") should be selected if the person is a director who studied in the US and the award is not of Festival-type.

An example rule from the `ripper` approach can be seen in Figure 4. The rules themselves look interesting, but while they improve in precision, as was our goal, their lack of recall makes their current implementation unsuitable for use. We have identified a number of changes that we could make to improve this process and discuss them at the end of the next section. Given the experimental nature of these results, we would not yet draw any conclusions about the ultimate benefit of the `ripper` approach.

5 Related Work

Very few researchers have addressed the problem of knowledge acquisition for content selection in generation. A notable exception is Reiter et al. (2000)'s work, which discusses a rainbow of knowledge engineering techniques (including direct acquisition from experts, discussion groups, etc.). They also

Experiment	development				`imdb.com`			
	Selected	Prec.	Rec.	F*	Selected	Prec.	Rec.	F*
baseline	530	0.40	0.72	0.51	727	0.35	0.68	0.46
class-based	550	0.41	0.94	0.58	891	0.36	0.88	0.51
content-selection	336	0.46	0.53	0.49	375	0.44	0.44	0.44
test set	293	1.0	1.0	1.0	369	1.0	1.0	1.0
select-all	1129	0.26	1.00	0.41	1584	0.23	1.00	0.37

Table 1: Experiment results

mention analysis of target text, but they abandon it because it was impossible to know the actual criteria used to chose a piece of data. In contrast, in this paper, we show how the pairing of semantic input with target text in large quantities allows us to elicit statistical rules with such criteria.

Aside from that particular work, there seems to exist some momentum in the literature for a two-level Content Selection process (e.g., Sripada et al. (2001), Bontcheva and Wilks (2001), and Lester and Porter (1997)). For instance, distinguish two levels of content determination, "local" content determination is the *"selection of relatively small knowledge structures, each of which will be used to generate one or two sentences"* while "global" content determination is *"the process of deciding which of these structures to include in an explanation"*. Our technique, then, can be thought of as picking the global Content Selection items.

One of the most felicitous Content Selection algorithms proposed in the literature is the one used in the ILEX project (Cox et al., 1999), where the most prominent pieces of data are first chosen (by means of hardwired "importance" values on the input) and intermediate, coherence-related new ones are later added during planning. For example, a painting and the city where the painter was born may be worth mentioning. However, the painter should also be brought into the discussion for the sake of coherence.

Finally, while most classical approaches, exemplified by (McKeown, 1985; Moore and Paris, 1992) tend to perform the Content Selection task integrated with the document planning, recently, the interest in automatic, bottom-up content planners has put forth a simplified view where the information is entirely selected before the document structuring process begins (Marcu, 1997; Karamanis and Manurung, 2002). While this approach is less flexible, it has important ramifications for machine learning, as the resulting algorithm can be made simpler and more amenable to learning.

6 Conclusions and Further Work

We have presented a novel method for learning Content Selection rules, a task that is difficult to perform manually and must be repeated for each new domain. The experiments presented here use a resource of text and associated knowledge that we have produced from the Web. The size of the corpus and the methodology we have followed in its construction make it a major resource for learning in generation. Our methodology shows that data currently available on the Internet, for various domains, is readily useable for this purpose. Using our corpora, we have performed experimentation with three methods (exact matching, statistical selection and rule induction) to infer rules from **indirect** observations from the data.

Given the importance of content selection for the acceptance of generated text by the final user, it is clear that leaving out required information is an error that should be avoided. Thus, in evaluation, high recall is preferable to high precision. In that respect, our *class-based* statistically selected rules perform well. They achieve 94% recall in the best case, while filtering out half of the data in the input knowledge base. This significant reduction in data makes the task of developing further rules for content selection a more feasible task. It will aid the practitioner of NLG in the Content Selection task by reducing the set of data that will need to be examined manually (e.g., discussed with domains experts).

We find the results for `ripper` disappointing and think more experimentation is needed before discounting this approach. It seems to us `ripper` may be overwhelmed by too many features. Or, this may be the best possible result without incorporating domain knowledge explicitly. We would like to investigate the impact of additional sources of knowledge. These alternatives are discussed below.

In order to improve the rule induction results, we may use *spreading activation* starting from the particular frame to be considered for content selection and include the semantic information in the local context of the frame. For example, to content select ⟨`birth date year`⟩, only values from frames ⟨`birth date`⟩ and ⟨`birth`⟩ would be considered (e.g., ⟨`relative...`⟩ will be completely disregarded). Another improvement may come from more intertwining between the exact match and statistical selector techniques. Even if some data path appears to be copied verbatim most of the time, we can run our statistical selector for it and use held out data to decide which performs better.

Finally, we are interested in adding a domain paraphrasing dictionary to enrich the *exact matching* step. This could be obtained by running the semantic input through the lexical chooser of our biography generator, PROGENIE, currently under construction.

References

Liane Acker and Bruce W. Porter. 1994. Extracting viewpoints from knowledge bases. In *Nat. Conf. on Artificial Intelligence*.

Kalina Bontcheva and Yorick Wilks. 2001. Dealing with dependencies between content planning and surface realisation in a pipeline generation architecture. In *Proc. of IJCAI'2001*.

William Cohen. 1996. Learning trees and rules with set-valued features. In *Proc. 14th AAAI*.

Richard Cox, Mick O'Donnell, and Jon Oberlander. 1999. Dynamic versus static hypermedia in museum education: an evaluation of ILEX, the intelligent labelling explorer. In *Proc. of AI-ED99*.

Pablo A Duboue and Kathleen R McKeown. 2003a. ProGenIE: Biographical descriptions for intelligence analysis. In *Proc. 1st Symp. on Intelligence and Security Informatics*, volume 2665 of *Lecture Notes in Computer Science*, Tucson, AZ, June. Springer-Verlag.

Pablo A Duboue and Kathleen R McKeown. 2003b. Statistical acquisition of content selection rules for natural language generation. Technical report, Columbia University, Computer Science Department, June.

I. P. Fellegi and A. B. Sunter. 1969. A theory for record linkage. *Journal of the American Statistical Association*, 64:1183–1210, December.

Gabriel Illouz. 2000. *Typage de données textuelles et adaptation des traitements linguistiques, Application à l'annotation morpho-syntaxique*. Ph.D. thesis, Universit Paris-XI.

Nikiforos Karamanis and Hisar Maruli Manurung. 2002. Stochastic text structuring using the principle of continuity. In *Proc. of INLG-2002*.

S. Kim, H. Alani, W. Hall, P. Lewis, D. Millard, N. Shadbolt, and M. Weal. 2002. Artequakt: Generating tailored biographies with automatically annotated fragments from the web. In *Proc. of the Semantic Authoring, Annotation and Knowledge Markup Workshop in the 15th European Conf. on Artificial Intelligence*.

James Lester and Bruce Porter. 1997. Developing and empirically evaluating robust explanation generators: The knight experiments. *Comp. Ling.*

Daniel Marcu. 1997. From local to global coherence: A bottom-up approach to text planning. In *Proceedings of Fourteenth National Conference on Artificial Intelligence (AAAI-1997)*, pages 450–456.

Kathleen Rose McKeown. 1985. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge University Press, Cambridge, England.

Johanna D. Moore and Cecile L. Paris. 1992. Planning text for advisory dialogues: Capturing intentional and rhetorical information. *Comp. Ling.*

Dragomir Radev and Kathleen R. McKeown. 1997. Building a generation knowledge source using internet-accessible newswire. In *Proc. of the 5th ANLP*.

Ehud Reiter, R. Robertson, and Liesl Osman. 2000. Knowledge acquisition for natural language generation. In *Proc. of INLG-2000*.

Barry Schiffman, Inderjeet Mani, and Kristian J. Concepcion. 2001. Producing biographical summaries: Combining linguistic knowledge with corpus statistics. In *Proc. of ACL-EACL 2001*.

Sidney Siegel and John Castellan Jr. 1988. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, New York, 2nd edition.

Somayajulu G. Sripada, Ehud Reiter, Jim Hunter, and Jin Yu. 2001. A two-stage model for content determination. In *ACL-EWNLG'2001*.

Elke Teich and John A. Bateman. 1994. Towards an application of text generation in an integrated publication system. In *Proc. of 7th IWNLG*.