Input Seed Features for Guiding the Generation Process: A Statistical Approach for Spanish

Cristina Barros
Department of Software
and Computing Systems
University of Alicante
Apdo. de Correos 99
E-03080, Alicante, Spain
cbarros@dlsi.ua.es

Elena Lloret
Department of Software
and Computing Systems
University of Alicante
Apdo. de Correos 99
E-03080, Alicante, Spain
elloret@dlsi.ua.es

Abstract

In this paper we analyse a statistical approach for generating Spanish sentences focused on the surface realisation stage guided by an input seed feature. This seed feature can be anything such as a word, a phoneme, a sentiment, etc. Our approach attempts to maximise the appearance of words with that seed feature along the sentence. It follows three steps: first we train a language model over a corpus; then we obtain a bag of words having that concrete seed feature; and finally a sentence is generated based on both, the language model and the bag of words. Depending on the selected seed feature, this kind of sentences can be useful for a wide range of applications. In particular, we have focused our experiments on generating sentences in order to reinforce the phoneme pronunciation for dyslalia disorder. Automatic generated sentences have been evaluated manually obtaining good results in newly generated meaningful sentences.

1 Introduction

The task of Natural Language Generation (NLG) comprises a wide range of subtasks which extend from an action planning until its execution (Bateman and Zoch, 2003). This subtasks are commonly viewed as a pipeline of three stages: document planning, microplanning and surface realisation (Reiter and Dale, 2000).

The NLG can be applied to several fields, not only to the task of reporting, such as text simplification (Reiter et al., 2009), recommendation generation (Lim-Cheng et al., 2014), text summarisation (Portet et al., 2007) or text that attempts to help people having any kind of disorders in therapies (Black et al., 2012).

Despite the applicability of NLG, this is not a trivial task. There is still a lot of room for improvement, and small steps in this task would be useful for being integrated or applied in larger NLG or NLP systems.

Therefore, the main goal of this paper is to present and evaluate a statistical NLG approach for Spanish based on N-grams language models. Our approach is focused on the surface realisation stage, and it is initially designed and tested for Spanish, but it can be extrapolated to other languages as it is statistical-based. The novelty of this approach lies in its input data, which can be a concrete seed feature or aspect (communicative goal) that we will be used to guide the generation of the new sentence (i.e., for guiding the generation process). This seed feature could be a word, a phoneme, a sentiment, etc.

This type of generated sentences can be useful in many different ways such as helping in therapies as has been outlined above. Specifically, we have chosen stories generation as our experimental scenario, so that a person with dyslalia, a speech disorder that implies the inability of pronounce certain phonemes, can reinforce the pronunciation of several problematic phonemes through reading and repeating words. So the aim of these sentences for dyslalia would be to contain a huge number of words with a concrete phoneme.

At this stage we are not exhaustively evaluating how syntactically and semantically correct a sentence is, but just whether to what extent a sentence fulfilling a communicative goal can be generated from a functional point of view. We consider that the communicative goal of our experimental scenario is to teach how a phoneme should be pronounced, so, by repeating the desired phoneme along a sentence this goal can be reached. Therefore, we will evaluate and analyse the output from our approach based on the seed feature appearance along the sentence and the sentence correctness.

The remainder of this paper is as follows. Section 2 discusses some related work concerned with surface realisation statistical systems. Section 3 presents our statistical approach for NLG based on seed features. Section 4 shows the experimentation carried out over the approach. In Section 5 the evaluation and the results obtained is discussed. Section 6 presents the potentials and limitations of our approach. Finally, section 7 draws some conclusions and outlines ideas for future work.

2 Related Work

The use of statistical techniques in NLG have been widely spread since Langkilde and Knight (1998) used them for the first time, where they used language models (LM) to choose words transformations after applying generation rules. Most of these techniques use language models, such as n-grams, or stochastic grammars. An example of these statistical techniques are given in (Kondadadi et al., 2013) that presents a statistical NLG system which consolidates macro and micro planning, as well as surface realisation stages into one statistical learning process. Moreover, many other statistical examples can be found in (Lemon, 2008), where a new model for adaptive NLG in dialog, showing how NLG problems can be approached as statistical planning problems using reinforcement learning, is presented. In the BAGEL system (Mairesse et al., 2010), a statistical language generator which uses dynamic Bayesian networks to learn from semantically-aligned data is integrated.

These statistical LM have been employed with several languages including Chinese, English, German and Spanish (Bohnet et al., 2010), where they take advantage of multilevel annotated corpora and propose a multilingual deep stochastic sentence realiser.

On the other hand, regarding to the application of NLG in order to help people having any type of problem or disorder there are several systems. For instance, STOP (Reiter et al., 2003) that generates letters to dissuade users from smoking, or systems to reduce anxiety of patients with cancer by providing them with information (Cawsey et al., 2000). These two systems employ templates that are filled with information from a data base or a knowledge base selected from user profiles.

There are approaches, such as the one in (Fernández et al., 2008) that generates sentences in Spanish containing words related to a specific restricted scenario, but, to the best of our knowledge, there is not a research in NLG focused on generating sentences in Spanish with the restriction of containing words with a specific seed feature. Moreover, since we use probabilistic techniques, these are language independent allowing its application to others languages adapting the necessary resources (e.g., semantic features) for the language-specific part.

3 Our Seed Feature Guided Language Generation Process

We propose a statistical approach using n-gram LM guided by an input seed feature. This approach is focused on generating a sentence with the highest number of words containing a certain seed feature. This seed feature, used to guide all the generation process, can be anything, such as letters, phonemes, POS tag, sentiments, etc.
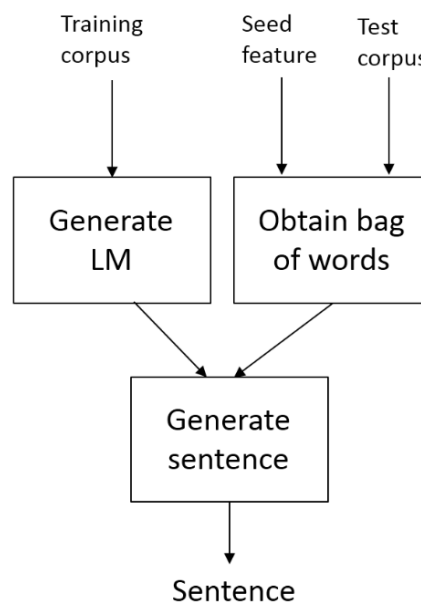

Figure 1: Our approach diagram

The input of this approach are: i) a training corpus, ii) a test corpus and iii) the seed feature. In Figure 1 a diagram of the process flow can be seen.

In the following paragraphs it is explained how the approach works.

1. **Generate the language model**: Before starting with the process, we train the LM over a training corpus in the desired language.

2. **Obtain the bag of words**: We obtain from the test corpus a bag of words having the seed

10

feature which is going to be used for the generation. This bag of words includes the word itself and its frequency of occurrence in the test corpus.

3. **Generate the sentence**:

This step of the process can be executed with two different configurations. The default configuration only generates one sentence based on the seed feature; and, with the overgeneration configuration, the system generates several sentences based on the seed feature. Next, we will explain the overall functioning of the process.

The approach is an iterative process in which this stage is repeated until either the desired length, or the special token end of sentence (</s>) are reached.

Assuming that there is a word that has been obtained from the previous iteration, we first search in the bag of words if there is a word in it that follows the word from the previous iteration. If so, we check which one has the highest probability based on the LM depending on that word, and in case of a draw between two or more words, then the word chosen is the one with a higher frequency in the test corpus.

Otherwise, we look for the word which has the highest probability of appearance with the word selected from the previous iteration in our LM, and if there are more than one word in our LM with the same probability, we check if any of them contains the seed feature. In that case, we pick the word with the seed feature; in another case, we choose the first appearance of the word with highest probability. As we said before, the process runs, prioritising the selection of words containing the seed feature, until the desired length or the token (</s>) are reached.

We took several issues into consideration during the implementation of our approach. For the first iteration, we initially set the special token start of the sentence (<s>) as our starting word. Moreover, when we choose the words it is taken into account that, if the chosen word is a stopword, then the process returns the stopword accompanied with the most probable next word. Another issue taken into account is that a stopword is not selected as the next word on the last iteration,

to prevent sentences ending inappropriately. Finally, a word cannot be chosen if it has been chosen before. This is to avoid words or word's sequences repetitions along the sentence.

The main difference between our two configurations lies on the first iteration of the generation process. With the default configuration, we only choose one initial word, so a single sentence is generated. With the overgeneration configuration, for an input seed feature, a list of words is chosen. This list contains the words that i) have the same probability as the one with the highest probability of appearance with the token (<s>), and ii) are within a range of less than a 0.5% of probability with respect to the words with the highest probability of appearance with the token (<s>) (this was empirically determined). In the remainder iterations, for each word contained in the list, the process runs likewise the default configuration.

4 Experimental setup

In this section we are going to discuss both the scenario, resources and tests performed to the approach.

4.1 Scenario

We place our research in the context of generating text to help people with a any kind of disorder. In particular, generating stories in order to help children with dyslalia could be one of the applications encapsulated within this application area (Barros and LLoret, 2015). Dyslalia is a disorder in phoneme articulation which implies the inability to correctly pronounce certain phonemes or groups of phonemes (in Spanish some of this phonemes are: /ch/, /ll/, /rr/ or in English are: /zh/, /ng/, /j/). This disorder is estimated to have a 5-10% incidence among the child population (Conde-Guzón et al., 2014).

Consequently, and based on the dyslalia disorder, the seed feature selected in order to generate the sentences is a problematic phoneme. Therefore, our main objective is to generate Spanish sentences containing a large number of words with a concrete phoneme, so that a child with dyslalia can reinforce the phoneme's pronunciation through reading and repeating words. In Figure 2 an illustrative example in Spanish for the phoneme

/a/ obtained from a real story[1], being part of an educational project of the Spanish Government, can be seen. This type of sentences can be useful for dyslalia disorder because they reinforce the phoneme pronunciation of the child by constantly repeating that concrete phoneme.
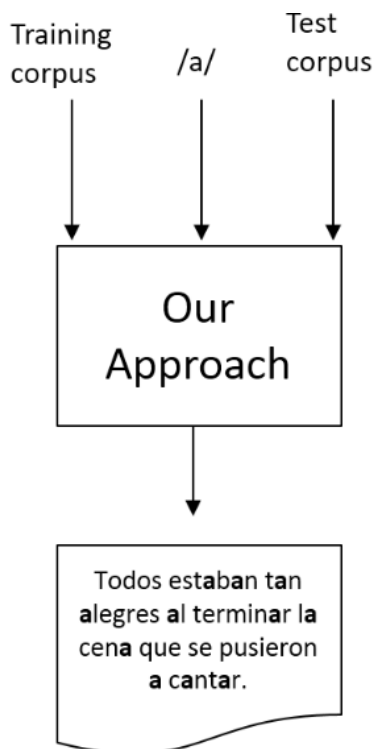

Figure 2: Illustrative example sentence. (*Translation: Everyone was so happy after dinner that began to sing.*)

4.2 Corpus and Resources

Since, as seen in the previous section, the scenario proposed is focused on generating stories which would improve the pronunciation of phonemes in children with dyslalia, the chosen corpus selected to perform the test is a collection of Hans Christian Andersen[2] stories in Spanish.

This collection consists of 158 children's stories (containing 21,085 sentences in total) of which 25% has been used as the test corpus from where the bag of words is obtained. For training the LM we have used the 75% of the corpus, in our case, we have trained a bigram LM and a trigram LM, being these models the most commonly used in

n-gram language model (Rosenfeld, 2000). If we had chosen any higher n, we will have to confront with data sparseness problems, where most possible grammatical n-grams would never appear even in huge training corpora.

These LMs have been trained using the SRILM (Stolcke, 2002) software that is a toolkit for building and applying statistical language models. We have chosen this software for its usability and because factored languages models (Bilmes and Kirchhoff, 2003) are implemented in it, and, in the future, we want to introduce them to the approach.

Obtaining words containing a concrete phoneme was performed according to the correspondence between phonemes and letters, employing some of the phonetic restrictions exposed in Morales (1992).

Furthermore, the stopword's file used in the experimentation has been obtained from the NLTK software data[3].

4.3 Experiments

We have performed several experiments dividing them in two groups that will be explained in more detail in the following paragraphs:

- Preliminary experiments

- Overgeneration experiments

To determine the length of the sentences to be generated, the average sentence length of the corpus was computed (16 words), using also this value for our experiments.

4.3.1 Preliminary experiments.

This type of experiments were conducted in order to check if it was worthy to carry on with this statistical-based approach, employing bigrams and trigrams LM, and to what extent the approach's behavior could be affected by the inclusion (or not) of stopwords. In addition, these experiments were carried out with the default configuration of the approach and testing all the Spanish phonemes. In this sense, we performed three types of experiments:

- First experiment: we removed the stopwords from the generation approach but we did not remove them from the training corpus.

Phoneme: ñ **Letters:** ñ

1st experiment.
Bigram: <s> añadió niño pequeño ruiseñor </s> (*said little boy nightingale*)
Trigram: <s> añadió aguja niños pequeño ruiseñor siguió cantando acompañado soñar mañana sueños </s>
(*said needle kids little nightingale continued singing accompanied morning dreaming dreams*)

2nd experiment.
Bigram: <s> mañana pequeña </s> (*little morning*)
Trigram: <s> mañana siguiente niño pequeño ruiseñor siguió cantando acompañado novia domesticada </s>
(*next morning little boy nightingale sung accompanied domesticated girlfriend*)

3rd experiment.
Bigram: <s> añadió niño pequeño ruiseñor </s> (*said little boy nightingale*)
Trigram: <s> añadió la aguja </s> (*said the needle*)

Figure 3: Preliminary experiments output[4]

- Second experiment: we trained both LMs without stopwords, and consequently the generation was made without stopwords.

- Third experiment: we trained both LMs with stopwords and we also removed the words repetitions on the final sentence. Furthermore, the stopwords were included in the final sentence.

4.3.2 Overgeneration experiments.

This experiment was performed after checking the results from the preliminary experiments. The main objective of this experiment was to test the overgeneration configuration of the approach with all the Spanish phonemes, and, check if it generates some meaningful sentences, as well as the most common types of errors.

5 Evaluation and Discussion

In this section we report the results from our two types of experiments: preliminary and overgeneration experiments. Furthermore, for the resulting generated sentences we made a manual analysis and evaluation. With this evaluation we needed to check if there was any meaningful sentence, ensuring that the sentence contained at least one word with the concrete phoneme.

5.1 Preliminary experiments evaluation.

As previously explained in section 4.3.1, within these preliminary experiments we performed three types of tests regarding the approach behavior and the utilisation or not of stopwords. Some sentences obtained from this test can be seen in the Figure 3.

Concerning our first experiment, in many cases the approach did not find the next word and the generation ended before reaching the limit length of the sentence using both LMs, bigram and trigram. This was due to the fact that there are verbs or words that only appears next to stopwords. We also tested in this experiment that, when a stopword was found, the next function word returned the stopword accompanied with its next word, but the stopword was not included in the final sentence and it was only used for the next word prediction. Yet still, most generated sentence were meaningless and presented quite a lot repetition.

As a result of our second experiment, the generated sentences tend to be a sequence of nouns, verbs and adjectives without any relation between them.

Finally, in our third experiment we observed that the generated sentences with trigrams ended with the special token end of sentence (</s>), containing at least one word with the phoneme, and some of them where meaningful sentences. Regarding the bigrams generated ones, most of them contained a huge number of words with the phoneme but the words itself did not have any connection with each other.

Thanks to these results we found that our approach worked well in some cases and because of that we decided to try the overgeneration configuration of the approach.

Sentences	Local percentage (based on 95 sent.)	Global percentage (based on 208 sent.)
Generated sent. from bigram LM with (</s>)	46.32%	21.15%
Generated sent. from trigram LM with (</s>)	78.95%	36.06%
Newly generated not included in the corpus	73.68%	33.65%
Meaningful total sentences	56.84%	25.96%
Meaningful sentences included in the corpus	25.26%	11.54%
Newly meaningful generated sent. not included in the corpus	31.58%	14.42%
Newly meaningful generated sent. from bigram LM	9.47%	4.33%
Newly meaningful generated sent. from trigram LM	22.11%	10.10%

Table 2: Statistics of the generated sentences ended with (</s>)

5.2 Overgeneration experiments evaluation.

Based on the results of the preliminary experiments, we further test the overgeneration configuration (section 4.3.2).

In this case, the approach generated 208 sentences, which 119 of them contains the special token end of sentence (</s>). All these sentences were generated from the bigram and trigram LMs, so it can occur that the same sentence could be generated by both LMs. These sentences ended with the token (</s>) are important because they can be comparable to a complete sentence. Of the 119 sentences generated containing the token (</s>), 95 of them are different. This can be seen on Table 1.

Sentences	Number of generated sentences	Percentage
Total	208	100%
Not ended with (</s>)	89	42.79%
Ended with (</s>)	119	57.21%
Ended with (</s>) without repetition	95	45.67%

Table 1: Statistics of the generated sentences from the overgeneration configuration

We then focused the evaluation and analysis of our results on the sentences ending with the token (</s>). This is because we consider these sentences as complete sentences being this token comparable to a full stop. The statistics of Table 2 were calculated according to the total number of different generated sentences ended with the token (</s>), 95 sentences. In this Table we also include the comparative percentage regarding the total number of generated sentences, that is 208 sentences. As we can see in this Table, the statistics of meaningful sentences are really encouraging.

These meaningful sentences do not include punctuation marks so, although some sentences at first glance do not seem coherent, with the inclusion of some punctuation marks they become meaningful.

<s> allí se quedó con la doncella había llegado el invierno </s>
  (P: /a/ L: a T:*stayed there with the maid had come winter*)
<s> añadió el niño pequeño ruiseñor </s>
  (P: /ñ/ L: ñ T: *said the little boy nightingale*)
<s> dónde está el cielo </s>
  (P: /n/ L: n T: *where is the sky*)
<s> finalmente llegaron a la superficie del agua </s>
  (P: /f/ L: f T: *finally they reached the surface of the water*)
<s> nadie conoce la princesa </s>
  (P: /n/ L: n T: *nobody knows the princess*)
<s> pues bien hecho está </s>
  (P: /e/ L: e T: *well done*)
<s> quién pudiera verlo </s>
  (P: /u/ L: u T: *who could see*)
<s> verdad que es la vida </s>
  (P: /b/ L: b,v,w T: *really that is the life*)
<s> verdad vieja bruja perversa </s>
  (P: /b/ L: b,v,w T: *really wicked old witch*)

Figure 4: Newly meaningful generated sentences. P: phoneme, L: letters and T: translation in brackets

Meaningful sentences cover almost the half of the different sentences ended with the token (</s>), and those newly sentences that not explicitly exist on the training corpus are about 30% of

Error types		Number of sentences	Local percentage (95 sent.)
Grammatical concordance	Nominal	2	4.88%
	Verbal	7	17.07%
Non words semantic relations		36	87.80%
Missing main verb		7	17.07%
Incorrect syntactic order		38	92.68%

Table 3: Common types of generated sentences errors

the 95 different sentences with the token. These result are quite positive considering that we are only focusing on the appearance of words with the phoneme within the sentence. Moreover, trigram LM is more suitable than bigram LM since it generates a higher number of newly meaningful sentences. Some of these newly generated sentences, that have been created employing different phonemes, can be seen in the Figure 4.

5.2.1 Error Types and Analysis

After analysing the generated sentences ending with the special token end of sentence (</s>), they may have some common errors along the meaningless sentences. These errors affects the coherence and cohesion of the sentence leading to make the sentence meaningless.

We manually analysed all the generated sentences and classified these errors attending to frequent grammatical errors[5] and frequent drafting errors[6]. In this classification we do not take into account punctuation marks errors because, when we train the language models we remove the punctuation marks from the corpus and, consequently, when we generate the sentences we do not introduce them.

We have found morphosyntactic errors of concordance. We subdivided concordance into two levels: nominal and verbal. Errors in nominal concordance refers to errors regarding with gender and number of the words, and, on the other hand, errors in verbal concordance refers to discordance between verb and subject in number. We also found errors regarding semantics relation between words, that is, the meaning of the words are unrelated to each other. Furthermore, we also found sentences not having a main verb conjugated. The

most common error was found in the order of the words, having an incorrect order leading to nonsense sequences of words.

Because not all the sentences presents only one type of errors, in Table 3 we have counted each type of error independently, for the meaningless sentences ended with (</s>) that have that error. As it can be seen in the table, and it was already noted, the most common errors among the sentences are syntactic errors and non semantic relation between words. We will see some examples below with its translation in brackets. For example, a sentence with a missing main verb conjugated is:

<s>ahora **hacer** </s>(*now do*)

An example sentence having only nominal concordance error is:

<s>aquel**lo** era demasia**do** fina (</s>) (*this was too thin*)

Where *aquello* and *demasiado* are masculine and *fina* is feminine. And finally, an example sentence with ordering errors, non semantic words relations and verb concordance error:

<s>allí orgullo y aquella belleza brillan en aquellos pajarillos de ello </s>(*there pride and beauty that shine in those birds of it*)

These errors can be corrected employing grammars for generating a sentences with a correct syntactic order or using some kind of semantic information in order to select words related semantically to one another.

6 Potentials and Limitations of the approach

Considering this approach as our research starting point we have to take into account some key as-

---

[5]https://ciervalengua.
files.wordpress.com/2011/11/
errores-gramaticales-frecuentes.pdf

[6]http://blog.pucp.edu.pe/blog/
blogderedaccion/2013/04/18/
errores-m-s-comunes-en-la-redacci-n/

pects from where we can improve the approach until we can achieve a fully correct generation of correct syntactic and semantically generated sentences based on a seed feature.

This approach has a great potential since it is a statistical approach, that means that these type of techniques are language independent, so we only need to adapt the language-specific approach's input, resulting this adaptation cost not really high. Moreover, an advantage of our approach is that we can make a more flexible generation adapted to different scenarios and applications guided for the input seed feature, being our surface realisation approach flexible and adaptive.

There is much information to consider in order to form a correct sentence. On the one hand we need syntactic information in order to get a correct syntactic structure of the sentence. This syntactic structure information can be achieved via grammars or trees. We will check existing Spanish resources in order to decide if we use one of them or develop our own. For the other hand, we need semantic information to make the generated sentences coherent. There are several linguistic theories that refers to discourse coherence such as the rhetorical structure theory (Mann and Thompson, 1988) or the systemic functional linguistic (Matthiessen and Halliday, 1997), that could be further exploited and integrated in the approach.

7 Conclusions and Future Work

We have presented a statistical NLG approach for Spanish guided by a seed feature. This approach allow us to create sentences containing a large number of words with a concrete seed feature. We also outlined a possible NLG scenario where these sentences can be helpful in speech therapies. For example, if the selected seed feature is a phoneme, this kind of sentences can be used in order to improve phoneme pronunciation.

Furthermore, we have shown that the approach obtains good results generating meaningful sentences not contained in the training corpus, taking into account that we are only focusing on the appearance of words with the concrete seed feature. Although the results obtained are promising, we must improve them because we do not generate meaningful sentences in all cases.

In the future, the approach will be modified to include both semantic and syntactic information to ensure that the generated sentences will be syntactic and semantically correct. We also want to test and subsequently include to the approach factored language models. In this model enunciated by Bilmes and Kirchhoff (2003) a word is viewed as a vector of factors that can be anything, including morphological classes, stems, roots, semantic information, etc. The main goal of this model is to produce a language model taking into account these factors. So, this type of model can serve us as a way of combine different information at a word level with our seed feature-based approach. In addition, once we have consolidated this model with our approach, we will test it with an English corpus in order to compare its results with the ones obtained from employing a Spanish corpus.

Finally, we need to investigate diverse ways of evaluating the generated sentences instead of manually evaluate this sentences. This will allow us in the future to have an homogeneous way of evaluating the generated sentences from an objective point of view.

References

Cristina Barros and Elena LLoret. 2015. Proposal of a data-to-text natural language generation approach to create stories for dyslalic children. In *1st International Workshop on Data-to-text Generation*, Edinburgh.

John Bateman and Michael Zoch. 2003. *Natural Language Generation*. Oxford University Press.

Jeff A. Bilmes and Katrin Kirchhoff. 2003. Factored language models and generalized parallel backoff. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 4–6. Association for Computational Linguistics.

Rolf Black, Annalu Waller, Ross Turner, and Ehud Reiter. 2012. Supporting personal narrative for children with complex communication needs. *ACM Trans. Comput.-Hum. Interact.*, 19(2):15:1–15:35.

Bernd Bohnet, Leo Wanner, Simon Mille, and Alicia Burga. 2010. Broad coverage multilingual deep sentence generation with a stochastic multi-level realizer. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 98–106. Association for Computational Linguistics.

Alison J. Cawsey, Ray B. Jones, and Janne Pearson. 2000. The evaluation of a personalised health information system for patients with cancer. *User Modeling and User-Adapted Interaction*, 10(1):47–72.

Pablo Conde-Guzón, Pilar Quirós-Expósito, María Jesús Conde-Guzón, and María Teresa Bartolomé-Albistegui. 2014. Perfil neuropsicológico de niños con dislalias: alteraciones mnésicas y atencionales. *Anales de Psicología*, 30:1105 – 1114.

Carles Fernández, Xavier Roca, and Jordi Gonzàlez. 2008. Providing automatic multilingual text generation to artificial cognitive systems. *Vigo International Journal of Applied Linguistics*, 5:37–62.

Ravi Kondadadi, Blake Howald, and Frank Schilder. 2013. A statistical nlg framework for aggregated planning and realization. pages 1406–1415. The Association for Computer Linguistics.

Irene Langkilde and Kevin Knight. 1998. Generation that exploits corpus-based statistical knowledge. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pages 704–710. Association for Computational Linguistics.

Oliver Lemon. 2008. Adaptive natural language generation in dialogue using reinforcement learning. *Proc. SEM-dial*, pages 141–148.

Natalie R. Lim-Cheng, Gabriel Isidro G. Fabia, Marco Emil G. Quebral, and Miguelito T. Yu. 2014. Shed: An online diet counselling system. In *DLSU Research Congress 2014*.

François Mairesse, Milica Gašić, Filip Jurčíček, Simon Keizer, Blaise Thomson, Kai Yu, and Steve Young. 2010. Phrase-based statistical language generation using graphical models and active learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1552–1561. Association for Computational Linguistics.

William C. Mann and Sandra A. Thompson. 1988. Rhetorical structure theory: Toward a functional theory of text organization. *Text - Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281.

C.M.I.M. Matthiessen and M.A.K. Halliday. 1997. *Systemic Functional Grammar: A First Step Into the Theory*.

Juan Luis Onieva Morales. 1992. *Nuevo método de ortografía*. Colección Cervantes. Verbum.

François Portet, Ehud Reiter, Jim Hunter, and Somayajulu Sripada. 2007. Automatic generation of textual summaries from neonatal intensive care data. In *Artificial Intelligence in Medicine*, pages 227–236. Springer.

Ehud Reiter and Robert Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.

Ehud Reiter, Roma Robertson, and Liesl M Osman. 2003. Lessons from a failure: Generating tailored smoking cessation letters. *Artificial Intelligence*, 144(1):41–58.

Ehud Reiter, Ross Turner, Norman Alm, Rolf Black, Martin Dempster, and Annalu Waller. 2009. Using NLG to help language-impaired users tell stories and participate in social dialogues. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 1–8. Association for Computational Linguistics.

Ronald Rosenfeld. 2000. Two decades of statistical language modeling: Where do we go from here? In *Proceedings of the IEEE*, pages 1270–1278.

Andreas Stolcke. 2002. Srilm - an extensible language modeling toolkit. In *Proceedings International Conference on Spoken Language Processing, vol 2.*, pages 901–904.