HanaNLG: A Flexible Hybrid Approach for Natural Language Generation

Cristina Barros[(✉)] and Elena Lloret

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

Abstract. Nowadays, with advances in digital technologies, interaction between computers and humans is essential. In this regard, the area of Natural Language Generation (NLG) can provide techniques capable of facilitating and improving this type of interaction. However, the existing approaches to this field are usually developed ad-hoc for specific tasks, purposes and domains, which hinders the advancement of flexible and adaptable multi-domain NLG systems. Under these premises, the objective of this paper is to present HanaNLG, a hybrid generic NLG approach, focused on the surface realisation stage. HanaNLG combines statistic and knowledge-based techniques and is able to generate text independently of the domain. In particular, this is done by exploiting language models in conjunction with semantic knowledge, providing flexibility to the whole generation process, thus, minimising the high cost associated with the development of common elements involved in NLG, such as grammars. Therefore, taking into account this joint perspective, our approach contributes to advancing the NLG field by providing greater flexibility when it comes to (i) producing text for different domains, and (ii) increasing the variety of vocabulary to appear in the generated text. In order to assess the effectiveness of HanaNLG, it was tested in two domains: (i) NLG for assistive technologies and, (ii) NLG for creating opinionated sentences. The positive results obtained (almost the 99% of the generated sentences for both domains are original and well constructed) show that our approach is capable of generating text for different domains. More importantly, the combination of language models with semantic knowledge enhances the quality of the generated text, thereby improving the results obtained compared to other methods that only rely on statistical methods.

Keywords: Language generation · Opinionated sentences · VerbNet · WordNet

1 Introduction

In today's society, as technology advances, the importance of facilitating the interaction between humans and machines is evident. In this context, both the understanding and the production of language are essential factors for enabling

this interaction, but the flexibility and ambiguity of the natural language are unavoidable for a computer.

In this sense, Natural Language Processing (NLP) – a theory-motivated range of computational techniques for the automatic analysis and representation of human languages [1]— is an important and developing field. Natural Language Generation (NLG), a subfield of NLP, aims to automatically generate text by involving a wide range of subtasks that are usually grouped into a three stage pipeline [2]: macroplanning, microplanning and surface realisation. While the first two stages are focused on what content the generated text must include and how it should be structured, the final stage is responsible for generating the final output in a NLG system. In particular, the surface realisation stage has been commonly addressed from two distinct, but not mutually exclusive perspectives: i) knowledge-based techniques, that rely on linguistic theories to generate sentences; and, ii) statistical approaches, whose underlying idea is based on the analysis and computation of the probability of certain words appearing together and thereafter, studying the creation of a sentence given an initial set of words. Whereas knowledge-based approaches must be designed and developed for each specific domain, purpose or language, statistical approaches are more flexible in terms of these previous issues, but they can suffer from the lack of deep linguistic information when generating a text. Besides considering the type of approach to use when developing a NLG system, the task entails additional difficulties. Existing systems are usually created ad-hoc, lacking the necessary flexibility to adapt them to other domains or purposes. Therefore, the development of open and flexible approaches remains a challenge for the research community. In this regard, the combination of the two aforementioned approaches —knowledge-based and statistical— would allow the creation of more flexible NLG systems, favouring the independence of domain and purpose.

Considering this existing open challenge, the main objective of this paper is to present HanaNLG (Hybrid surfAce realisatioN Approach for Natural Language Generation), a hybrid generic approach for NLG focused on the surface realisation stage and which is capable of generating text regardless of its domain. HanaNLG is hybrid because it relies on the use of statistical information and semantic knowledge for producing the text. This type of hybrid approach in conjunction with seed features (i.e. abstract objects that will guide the generation with respect to its vocabulary) provide the flexibility to produce text for different domains and purposes. Proposing a flexible hybrid generic approach for NLG which combines statistical information with semantic knowledge for generating text enables the following contributions to the field: (i) text for different domains can be easily produced, and (ii) the variety of vocabulary that appears in the generated text is increased and improved.

The paper is structured as described next. Section 2 outlines the related work, focusing on hybrid approaches for NLG. Then, HanaNLG is described in Sect. 3. In Sect. 4, the experimentation environment, as well as the tools employed are explained. Section 5, presents and discusses the results. Finally, in Sect. 6, the main conclusions and directions for future work are provided.

2   Related Work

As previously mentioned, the task of NLG has been usually addressed from two different perspectives: knowledge-based [3,4] and statistical approaches [5,6]. The combination of these perspectives —which results in hybrid approaches— may overcome the flaws from each of them, leading to more flexible systems with respect to the domain, language or purpose.

From the end of the 20th century until today, hybrid approaches have been proposed to address the NLG task. One of the first approaches under this perspective is FERGUS (Flexible Rationalist-Empiricist Generation Using Syntax) [7]. This system addressed the microplanning and macroplanning stage with a combination of N-grams and tree-based statistical models, making use of a lexicalised tree-based syntactic grammar based on XTAG grammar [8]. The system FLIGHTS [9] is another example of a system that was developed to generate flight information adapted to the final user. In order to do that, FLIGHTS considered different knowledge bases, such as user models or dialog records, for selecting the content of the final output. Then, the text was realised employing the OpenCCG framework[1]. Recently, [10] presented a hybrid system, which first derived a template bank from a corpus and afterwards, selected the best template using a statistical ranking model. In [11], a multilingual approach for abstractive summarisation employing semantic representations was proposed. This approach, which underlying theoretical framework is the Meaning-Text Theory [12], relies on the use of statistical and rule-based techniques to produce a summary in response of a user query. In [13] a hybrid/symbolic approach that helps to model the constrains in the interactions between the different stages of a NLG system was proposed. To fulfil this purpose, the approach uses a small handwritten grammar, a statistical hypertagger and a surface realisation algorithm. Recently, [14] proposed a hybrid system for Spanish that generates sentences from pictograms. This is done by combining information from a lexicon and a language model to first infer prepositions and then, the sentences are generated using a self-constructed Spanish adaptation of SimpleNLG [15].

Finally, the work presented here differs from the existing hybrid approaches in that HanaNLG explicitly combines semantic knowledge with statistical information, in conjunction with seed feature to make the generation of text more flexible. In this manner, our approach can be easily adapted to be used for different domains and purposes.

3   HanaNLG: Our Proposed Approach

We propose HanaNLG, a hybrid generic approach focused on the surface realisation stage which combines the use of language models and semantic knowledge. In this regard, our approach employs Factored Language Models (FLM) as language models, and uses the semantic knowledge from linguistic resources, such as WordNet [16] and VerbNet [17]. The use of these type of techniques introduces

---

[1] http://openccg.sourceforge.net/.

flexibility to the whole generation process, allowing the production of text for different domains and purposes.
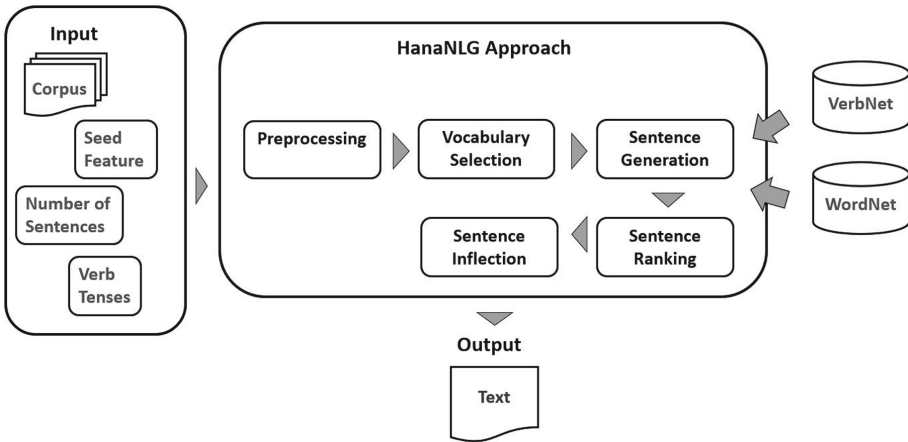

Fig. 1. Architecture of HanaNLG.

HanaNLG is based on over-generation and ranking techniques, where several sentences are first generated and are subsequently ranked, based on their probability, to only select the one with the highest probability. It is structured as a five-module architecture as depicted in Fig. 1. The input to the approach are: (i) a corpus, (ii) a seed feature, (iii) the number of sentences to generate and (iv) the verb tense for each of the sentences to be generated. In this sense, the corpus will be employed to obtain information about words, after preprocessing it; and it also will be used to train the FLM. In our approach, a seed feature is considered to be an abstract object (e.g. phonemes, sentiments, polarity, etc.) that will guide the generation process in terms of the vocabulary to be contained in the output text. The number of sentences indicates the total number of sentences that will be generated for the final output, and the verb tenses are used to specify the verb tense of each of the output sentences.

HanaNLG is capable of generating a complete text, where the sentences composing it are generated following the same strategy. Therefore, for each of the sentences to be generated, the tasks within this architecture that are performed as next summarised. In the following subsections, more details about them will be provided.

– **Preprocessing**: The input corpus is preprocessed with a linguistic analyser and is then tagged with different information. The corpus tagged is used to train the language models that will be used during the generation.
– **Vocabulary selection**: The vocabulary that will be employed in the generation process is selected based on the input seed feature from the tagged corpus. The vocabulary will be in lemma form.

– **Sentence generation**: Taking as input the vocabulary from the previous module, a set of lemmatised sentences is generated following an over-generation strategy using semantic linguistic resources as well as the language models.
– **Sentence ranking**: When a set of lemmatised sentences form is generated in the previous module, this module ranks them in order to select the one with the highest probability.
– **Sentence inflection**: Once a lemmatised sentence is selected in the previous module, is inflected according to the verb tense specified in the input.

3.1   Preprocessing

Before starting the generation process, taking a corpus as input to this stage, there are two main tasks to be performed. On the one hand, the corpus need to be linguistically analysed and tagged. On the other hand, the FLM needs the corpus to be tagged in order to be able to be trained. A language analyser tool is used to perform a linguistic analysis at different levels (e.g. lexical, syntactic and semantic). In this sense, information about the words themselves, their lemmas, their POS (Part-Of-Speech) tags and their synsets[2] is extracted. This information is then used to tag the entire corpus, resulting in a text similar to the one in Fig. 2.

Sentence raw: *Night fell and nothing...*

Sentence with format:
P-noun:X-NN:W-Night:L-night:S-noun|15167027_n
P-verb:X-VBD:W-fell:L-fall:S-verb|00151689_v
P-conjunction:X-CC:W-and:L-and:S-conjunction|none
P-pronoun:X-PRP:W-nothing:L-nothing:S-pronoun|none:N-none
...

Fig. 2. Example of the format of the corpus, where P: simple POS tag; X: full POS tag; W: word; L: lemma; and S: simple POS tag+synset

Once the corpus is tagged, the FLM can be trained. These models are an extension of the conventional language models where a word is viewed as a vector of k factors such that $w_t \equiv \{f_t^1, f_t^2, \ldots, f_t^K\}$. These factors can be anything, ranging from basic elements such as words to more complex elements such as rhetorical relationships. The main objective of these models is to create a probability language model over these factors. For this research work, the same linguistic information extracted from the corpus (i.e. words, lemmas, POS tags and synsets) is used as the factors to train the FLMs. These factors were selected due to the type of information they provide, since they provide us the necessary flexibility to adapt the vocabulary and the approach to different contexts.

---

[2] A set of synonyms used in Wordnet that are related to a term.

3.2   Vocabulary Selection

This second module of the architecture is in charge of selecting the content that will conform the final output of HanaNLG. In this respect, from the corpus tagged in the previous module, the words that are related to the input seed feature are selected. Depending on the seed feature, the detection of the words related to it may differ. For instance, it may not require the same type of resource the detection of a word associated with a negative polarity, as the detection of a word that may contain the phoneme /a/. Section 4 will describe the resources employed to identify the words related to the domains tested.
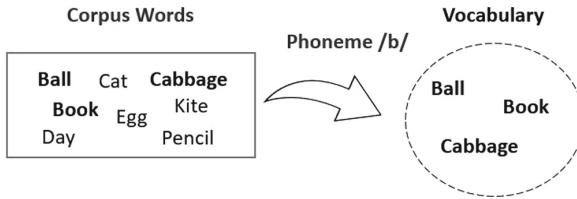
Fig. 3. Example of vocabulary selection for the phoneme /b/.

Once these words are selected, they are stored in a bag of words with their lemma form. This bag of words will be used during the generation process. An example of the vocabulary that would be selected for the phoneme /b/ is shown in Fig. 3.

3.3   Sentence Generation

Taking as input to this module the vocabulary previously gathered, the main objective of this module is to generate text maximising the words within this vocabulary. Since our approach relies on over-generation and ranking techniques, a set of sentences is generated to be later ranked in the next module. Each of these sentences is generated using the FLM previously trained in conjunction with VerbNet and WordNet resources. VerbNet is a verb lexicon for English that includes semantic as well as syntactic information about verbs. WordNet is a lexical-semantic database whose words are grouped into set of synonyms (i.e. synsets). From these computational linguistic resources, a set of syntactic and semantic frames is obtained and used as a basis for the generation. However, the type of information provided by both resources differs. On the one hand, the frames from VerbNet contain both, semantic and syntactic information about verbs. On the other hand, WordNet only provide a set of generic semantic frames for all the verbs included in its database. Figure 4 shows the different frames from VerbNet and WordNet obtained for the verb *to write*.

The sentences in this approach are generated from its core element, which we assume it is the verb. So, starting from a set of verbs from the vocabulary,

VERB: To Write

VerbNet frames	WordNet frames
NP V NP (Agent V Theme) NP V PP.topic (Agent V {about} Topic)	Somebody ----s something Somebody ----s something with something

Fig. 4. Example of the obtained frames for the verb *to write*.

their frames are extracted. In the case that the vocabulary does not contain any verb, a set of the most frequent verbs within the the input corpus is used instead. For each of these frames, a lemmatised sentence will be generated. These frames are first analysed to determine which elements of the sentence (i.e. the subject or the object of the sentence) are needed to be generated. For example, if a specific frame specifies that a *Subject* is needed, the elements of the subject will be generated based on the trained FLM, prioritising the words from the vocabulary. Likewise, if the *Object* is needed, it will be generated employing the same process. An example of the generated sentences for the frames in Fig. 4 is shown in Example 1.

(1) **NP V NP** → bob write a book.
NP V PP.topic → brooke write a book with a ballpoint pen.
Somebody —-s something → Bob write a card.
Somebody —-s something with something → Ben write a letter with a blue pen.

3.4 Sentence Ranking

Since HanaNLG follows a over-generation and ranking strategy, for each sentence that will form part of the final output, a set of lemmatised sentences is first generated and then ranked. Therefore the main objective of this module is to determine the one that will form part of this final output. In order to only select one sentence, a ranking based on the probability of the sentences is performed. The probability of a sentence is computed using the chain rule. In the chain rule, the probability of a sentence is calculated as the product of the probabilities of its words: $P(w_1, w_2...w_n) = \prod_{i=1}^{n} P(w_i|w_1, w_2...w_{i-1})$.

Depending on the language model used, the calculation of the probability of a word may differ. In our case, since we are using FLM, the probability of a word is calculated as a linear combination of FLM as suggested in [18], where a weight λ_i was assigned for each of them, being its total sum 1: $P(f_i|f_{i-2}^{i-1}) = \lambda_1 P_1(f_i|f_{i-2}^{i-1})^{1/n} + \cdots + \lambda_n P_n(f_i|f_{i-2}^{i-1})^{1/n}$, In this case, f corresponds to the selected factors to train the models.

The final lemmatised sentence selected in this ranking will be the one with the highest probability besides having the maximum number from the vocabulary (the ones that are related with the input seed feature). In Example 2 is show the sentence that will be selected from the ones in Example 1.

(2) bob write a book. $\rightarrow$ Probability: 0.25
brooke write a book with a ballpoint pen. $\rightarrow$ Probability: 0.40
Bob write a card. $\rightarrow$ Probability: 0.12
Ben write a letter with a blue pen. $\rightarrow$ Probability: 0.20

Selected Sentence: brooke write a book with a ballpoint pen.

3.5   Sentence Inflection

The last module of the architecture is the inflection one. The goal of this module is to inflect the words within the sentence selected by the previous module, based on the verb tense provided as input. As aforementioned, the words comprising the sentence are in lemma form; therefore, it is essential to inflect them to make the language as natural as possible.

In our approach, the inflection is addressed by the use of lexicons. These lexicons are used to obtain the desired inflection of the words. In addition to this, this module will make minor changes to the sentence regarding the concordance in number (singular and plural) and person (third person singular or others).

Once the sentence is inflected, it will form part of the final output of text generated by our approach. Example 3 of the final inflection, assuming that the verb tense is past simple, of the lemmatised sentence in Example 2.

(3) **Inflected Sentence:** Brooke wrote a book with a ballpoint pen.

4   Experiments

In order to demonstrate the flexibility of our HanaNLG, the experimentation was focused on the generation of text in English for two distinct scenarios that were proposed in [5]. On the one hand, (i) NLG for assistive technologies to reinforce people's pronunciation of specific phonemes and words; and, on the other hand, (ii) NLG for creating opinionated sentences to support users or systems with the generation of reviews and evaluative text. These scenarios will be described in the following subsections. We used a similar setting to the one in [5], in order to compare our results with theirs, since they employed a purely statistical approach for the generation.

In addition to these scenarios, during the development of HanaNLG, several tools were used. In this regard, the Freeling tool [19] was used to linguistically analyse the input corpus. For training the FLM, the SRILM software [20] which allows the building and training of several language models, was employed. In order to work with WordNet, the library JWI [21] was used, and in the case of VerbNet, the library JVerbnet[3] was employed. Concerning the ranking a linear combination of FLM was used to compute the probability of the words. In this sense, the FLM combination was as follows:

---

[3] http://projects.csail.mit.edu/jverbnet/.

$P(w_i) = \lambda_1 P(f_i|f_{i-2}, f_{i-1}) + \lambda_2 P(f_i|p_{i-2}, p_{i-1}) + \lambda_3 P(p_i|f_{i-2}, f_{i-1})$, where f refers to a lemma, p refers to a POS tag, and λ_i are set $\lambda_1 = 0.25$, $\lambda_2 = 0.25$ and $\lambda_3 = 0.5$. These values were empirically determined by testing different values and comparing the results obtained.

4.1 NLG for Assistive Technologies

Within the former, the experimentation was specifically focused on the generation of sentences for helping children with the dyslalia disorder. This disorder affects the articulation of phonemes, implying the inability to correctly pronounce certain phoneme or groups of phonemes. Based on this domain, the objective of the generated sentences is to have the maximum number of words with the problematic phoneme. This type of sentences has demonstrated to be useful in dyslalia speech therapies [22]. Subsequently, the seed feature in this domain is a phoneme. In this experiment, one sentence for each of the English phonemes (i.e. a total of 44 phonemes) was generated. The corpus employed in this case is a collection of 779 English children stories was used as corpora, including the Lobo and Matos corpus [23] and other stories automatically gathered from *Bedtime stories*[4] and *Hans Christian Andersen: Fairy Tales and Stories*[5].

4.2 NLG for Opinionated Sentences

Regarding our second scenario, the main objective of this experimentation was to generate sentences with a specific polarity (e.g. positive or negative) that will help users or systems in the creation of reviews and evaluative text. With the rise of Web pages where the people can express opinions through reviews or star ratings, the generation of these kind of sentences can be useful in the case of providing explanations or justifying the rating assigned to a product, movie, etc. In this case, the experimentation was focused on the context of movie reviews, generating a sentence for each of the polarities. For this purpose, the Sentiment Polarity Dataset [24] as our corpus for this domain.

5 Evaluation and Results

Evaluation in NLG is very challenging. Like in other research fields within NLP (e.g., summarisation or irony detection), there is a limited number of gold standards to compare the system output with. Moreover, if one wants to evaluate the meaning and structure of the generated text —as in the case of NLG— it becomes even more complex, since there are not automatic tools that allow us to do that. Therefore, the evaluation of NLG is usually performed manually and collaboratively [25]. Subsequently, in order to evaluate the sentences generated by our proposed approach, a user-based evaluation was conducted. A

---

[4] https://freestoriesforkids.com/.
[5] http://hca.gilead.org.il/.

total of 3 assessors participated in this manual evaluation, being these assessors graduate and post-graduate students from the Computer Science area with a proficient level of English. The assessors were asked to evaluate if the sentences were meaningful with respect to its coherence. The Kappa statistic [26] was used for computing the agreement between the assessors, obtaining a overall agreement of 0.84, which means that is a strong agreement between the assessors. In addition to this manual evaluation, we also measured the percentage of sentences that do not precisely appear in the training corpus, i.e., that were new and original with respect to the training corpora.

Table 1. Comparative results of the manual evaluation for the two domains proposed.

System	Domain	Meaningful sentences	Newly generated sentences
HanaNLG	NLG for assistive technologies	97.73%	100%
	NLG for opinionated sentences	100%	100%
Barros and	NLG for assistive technologies	95%	70%
Lloret 2017	NLG for opinionated sentences	100%	50%

Table 1 summarises the results obtained for the manual evaluation in comparison with the system presented in [5]. Barros and Lloret [5] presented a statistical surface realisation system with a similar experimentation setup, regarding the domains explored as aforementioned. However, their system only used statistical information for generating the sentences for both domains and the evaluation of the sentence was more permissive, being their generated sentences not inflected. In addition, their evaluation criteria of meaningfulness was different: if the sentence was meaningful by itself or the sentence become meaningful by adding prepositions or punctuation marks.

As can be seen in the table, HanaNLG outperforms the results obtained by Barros and Lloret. The sentences within our approach reach a 97.73% and a 100% of meaningful sentences for the NLG for assistive technologies and NLG for opinionated sentences, respectively. Regarding the creation of new content (i.e. the sentences generated that are not in the training corpus), HanaNLG generates utterly new sentences in contrast to Barros and Lloret system, were only 50% and 70% of their sentences are new.

In light of these results, it has been proved that the proposed hybrid perspective enhances the flexibility of the NLG process, being able to easily adapt the approach to generate text of different domains and purposes. Example 4 shows some examples of the generated sentences.

(4) **NLG for assistive technologies**
    Phoneme: /l/ **Sent:** *Hjalmar looked towards the sea.*
    Phoneme: /g/ **Sent:** *The hero fought with Argus.*

Phoneme: /th/ **Sent:** *The youth thanked God for the sooth.*
Phoneme: /d/ **Sent:** *The guard drew him through the field.*

NLG for opinionated sentences
Polarity: Positive **Sent:** *The apostle deserved the praise.*
Polarity: Negative **Sent:** *The jellyfish killed the member.*

6   Conclusions

This paper presented HanaNLG, a hybrid NLG approach focused on the surface realisation stage. The proposed approach uses statistical information from trained FLM in conjunction with semantic knowledge from several linguistic resources (i.e. VerbNet and WordNet) to generate text in a flexible manner. In order to assess this flexibility, our approach was tested in two different scenarios belonging to different domains and with different purposes: NLG for assistive technologies and NLG for opinionated sentences. On the one hand, in the first scenario, sentences that would be helpful in dyslalia speech therapies were produced. On the other hand, sentences with a specific polarity were generated for our second scenario.

A manual user-based and collaborative evaluation was conducted for evaluating the meaningfulness of the generated sentences in term of coherence. In addition to this, HanaNLG was compared to another state-of-the-art NLG approach that generated sentences for the same scenarios, with the difference that this approach is purely statistical, rather than hybrid like ours. The results obtained by our approach outperformed the ones from the state-of-the-art system, being almost the 99% (the overall of the two scenarios) of our generated sentences original (they do not appear in the training corpus) and meaningful.

In the future, we want to adapt HanaNLG to other languages. This could be done with the adaptation of the linguistic resources to languages other than English. In addition to this, we also want to explore other types of techniques, such as deep learning, to determine if the performance of our approach may improve. Furthermore, we also want to assess the generated sentence to verify if they really help to achieve the purpose that we have proposed in the scenarios.

References

1. Cambria, E., White, B.: Jumping NLP curves: a review of natural language processing research. Comput. Intell. Mag. IEEE **9**, 48–57 (2014)
2. Reiter, E., Dale, R.: Building Natural Language Generation Systems. Cambridge University Press, Cambridge (2000)

3. Wanner, L., Bohnet, B., Bouayad-Agha, N., Lareau, F., Nicklaß, D.: Marquis: generation of user-tailored multilingual air quality bulletins. Appl. Artif. Intell. **24**, 914–952 (2010)
4. McDonald, D.D.: 6. In: Natural Language Generation, pp. 121–144. CRC Press (2010)
5. Barros, C., Lloret, E.: A multilingual multi-domain data-to-text natural language generation approach. Procesamiento del Lenguaje Nat. **58**, 45–52 (2017)
6. Mairesse, F., Young, S.: Stochastic language generation in dialogue using factored language models. Comput. Linguist. **40**, 763–799 (2014)
7. Bangalore, S., Rambow, O.: Exploiting a probabilistic hierarchical model for generation. In: Proceedings of the 18th Conference on Computational Linguistics, vol. 1, COLING 2000, pp. 42–48. Association for Computational Linguistics (2000)
8. Group, X.R.: A lexicalized tree adjoining grammar for English. Technical report IRCS-01-03, IRCS, University of Pennsylvania (2001)
9. White, M.A.J., Clark, R.D., Moore, J.: Generating tailored comparative descriptions with contextually appropriate intonation. Comput. Linguist. **36**, 159–201 (2010)
10. Kondadadi, R., Howald, B., Schilder, F.: A statistical NLG framework for aggregated planning and realization. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, (Volume 1: Long Papers), pp. 1406–1415. Association for Computational Linguistics (2013)
11. Mille, S., Ballesteros, M., Burga, A., Casamayor, G., Wanner, L.: Multilingual natural language generation within abstractive summarization. In: Proceedings of the 1st International Workshop on Multimodal Media Data Analytics co-located with the 22nd European Conference on Artificial Intelligence, MMDA@ECAI 2016, pp. 33–38 (2016)
12. Žolkovskij, A.K., Mel'čuk, I.A.: O vozmožnom metode i instrumentax semantičeskogo sinteza. Naučno-techničeskaja informacija (1965)
13. Gardent, C., Perez-Beltrachini, L.: A statistical, grammar-based approach to microplanning. Comput. Linguist. **43**, 1–30 (2017)
14. García-Méndez, S., Fernández-Gavilanes, M., Costa-Montenegro, E., Juncal-Martínez, J., González-Castaño, F.J.: Automatic natural language generation applied to alternative and augmentative communication for online video content services using simpleNLG for Spanish. In: Proceedings of the Internet of Accessible Things. W4A 2018, pp. 19:1–19:4. ACM (2018)
15. Gatt, A., Reiter, E.: SimpleNLG: A realisation engine for practical applications. In: Proceedings of the 12th European Workshop on Natural Language Generation, pp. 90–93. Association for Computational Linguistics (2009)
16. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge (1998)
17. Schuler, K.K.: Verbnet: A Broad-coverage, Comprehensive Verb Lexicon. Ph.D. thesis (2005)
18. Isard, A., Brockmann, C., Oberlander, J.: Individuality and alignment in generated dialogues. In: Proceedings of the INLG, pp. 25–32. Association for Computational Linguistics (2006)
19. Padró, L., Stanilovsky, E.: Freeling 3.0: Towards wider multilinguality. In: Proceedings of the 8ht International Conference on Language Resources and Evaluation, European Language Resources Association (2012)
20. Stolcke, A.: SRILM - an extensible language modeling toolkit. Proc. Int. Conf. Spoken Lang. Process. **2**, 901–904 (2002)

21. Finlayson, M.A.: Java libraries for accessing the Princeton wordnet: comparison and evaluation. In: Proceedings of the 7th International Global WordNet Conference (GWC 2014), Tartu, Estonia, pp.78–85. Global WordNet Association (2014)
22. Rvachew, S., Rafaat, S., Martin, M.: Stimulability, speech perception skills, and the treatment of phonological disorders. Am. J. Speech-Lang. Pathol. **8**, 33–43 (1999)
23. Lobo, P.V., de Matos, D.M.: Fairy tale corpus organization using latent semantic mapping and an item-to-item top-n recommendation algorithm. In: Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010), European Languages Resources Association (ELRA) (2010)
24. Pang, B., Lee, L.: A sentimental education: Sentiment analysis using subjectivity. In: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, pp. 271–278. Association for Computational Linguistics (2004)
25. Gkatzia, D., Mahamood, S.: A snapshot of NLG evaluation practices 2005–2014. In: Proceedings of the 15th European Workshop on Natural Language Generation (ENLG), pp. 57–60. Association for Computational Linguistics (2015)
26. Randolph, J.J.: Online kappa calculator [computer software] (2008). http://justus.randolph.name/kappa