Research Progress on System-Generated Language: From the Psycholinguistic Perspective

Liming Zhao [a, b], Yufang Yang [a, *]
a, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
b, Graduate University of Chinese Academy of Sciences, Beijing, China
*Corresponding author. E-mail addresses: yangyf@psych.ac.cn

Abstract

As the development of computer technology and Internet use, system utterance producing gains more and more attention. Research on utterance producing systems has been approached from two angles. In one research tradition, the analysis of corpus led to templates of system utterance generation. In another tradition, a natural language generation (NLG) system corresponding to human language production theories was founded. The NLG system was marked as flexible and trainable. In this paper, we introduce the progress on system-generated language from a psycholinguistic perspective and take an example to explain the function of human language production theories for the development of NLG systems. As more and more progress was made by psycholinguists on language production, there will be abundant room for NLG systems to improve in the future.

Key Words: *system utterance, natural language generation, sentence planning, SPoT*

As the computer and Internet are used more and more widely and frequently, a lot of communication is replaced by human-computer interaction and chat online. The latter one is still a kind of communication between two human beings but with its own properties different from the face-to-face communication [1]. In this paper, we concern the system-generated language production during human-computer interactions. Since the technology of system utterance producing can save a lot of human resources, such as the operators, and make people get answers quickly through online enquiry, the development of utterance producing systems gain more and more attention in recent years. In the following part, we will generally introduce the different approaches to producing system utterances. Then we will focus on the natural language generation (NLG) system, which is flexible and trainable, and close to the human natural language production.

978-1-4244-6359-6/10/$26.00 ©2010 IEEE

1. Approaches to producing system utterances

We generate about 2 to 3 words per second on average in everyday conversation. It means we utter four syllables per second, or 10 or 12 phonemes. These words are sequentially selected from a large warehouse - Mental Lexicon. For adults with general reading and writing skills, this dictionary includes at least 5 to 10 million words. The words are produced rapidly and in highly complex way while the speakers make rare errors. In every 1,000 words produced, the errors on average is not more than 1 or 2 times [2]. It seems that we are born as conversationalists, but it is so difficult to teach computer to communicate as human beings. One of the key reasons is that the processing of speech production during human utterance is still puzzled. So in the past several years there have been a large increase in commercial spoken dialogue systems and template-based generation system. As many psycholinguists did lots of work on human cognition, and got achievements on speech production mechanism, there has been an increasing interest in the use of natural language generation in spoken dialogue.

1.1. Commercial spoken dialogue systems

The utterances in commercial spoken dialogue systems were highly scripted for style and register and recorded by voice talent. This is a traditional and simple technique for producing the system side of the conversation. Basing on the commercial need and the interoperability of technology vendors, platform integrators, application developers, and hosting companies, the commercial spoken dialog industry has reached a mature level [3].

The commercial dialogue systems aim at usability and task completion, so they are largely pragmatic in specific areas. However several factors argue against

the commercial dialog systems. The most salient one is that these spoken dialogue systems are lack of flexibility, especially compared to natural dialogue. It is difficult to apply a settled commercial dialogue system to another different domain. To be more flexible and support users widely, there should be some rules applied across different fields.

1.2. Template-based generation systems

In template-based generation, system utterances are produced from hand-crafted string templates with variables that are instantiated by the dialogue manager. The template-based generation was used widely in current research systems because it is conceptually easy to produce utterances across different domains, especially comparing with the commercial dialogue systems. Adapting an example from Reiter and Dale, a simple template-based system might start out from a semantic representation saying that the 306 train leaves Aberdeen at 10:00 am:

Departure ($train_{306}$, $location_{abdn}$, $time_{10:00}$),
and associate it directly with a template such as

[train] *is leaving* [town] *now* [4]

As we see, the templates are settled, so the variety of output is limited, as Reiter and Dale stated. They also pointed out that the template-based systems are more difficult to maintain and update [5]. At the same time, there should be lots of templates to be adjusted to the syntactic demands, such as subject–verb agreement and determiner–noun agreement. Busemann and Horacek made a tough comment that template-based systems do not embody generic linguistic insights [6], not to mention that the collection of templates becomes a software engineering problem as the complexity of the dialogue system increases. So there need to make the producing process more complex and more like the speech production system in human being.

1.3. Natural language generation (NLG) systems

From a technical perspective, almost all applied NLG systems divide the generation process into three modules:

Text Planning: Decide the conceptual goal of the utterances to be produced.

Sentence Planning: Use abstract linguistic formats to achieve the communicative goals.

Realization: Generate the individual sentences in a grammatically correct manner.

Such systems could start from the same input semantic representation as the template-based generation systems (e.g. the 306 train leaves Aberdeen at 10:00 am), subjecting it to a number of consecutive transformations until a surface structure being selected (e.g. this train is leaving Aberdeen now). The proposed architecture for a spoken dialogue system, including three modules of NLG, can be seen Figure 1.

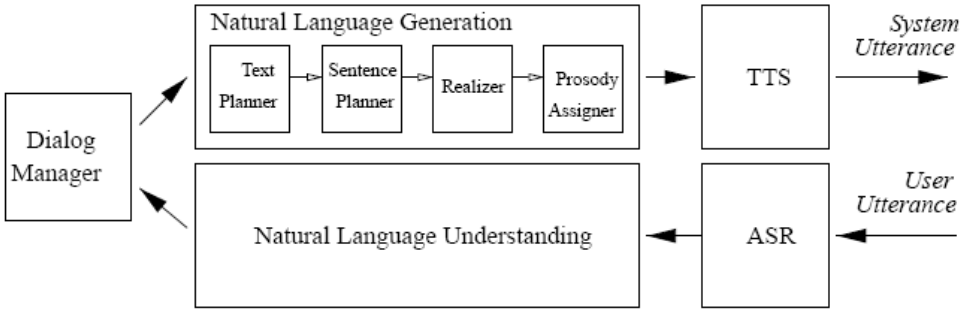

Figure 1. Architecture of a dialogue system with natural language generation [14]**.**

From the psycholinguistic perspective, the three modules of NLG systems just correspond to the three processing levels of language production. So far, most psycholinguistic models of language production have agreed that at least the following three processing levels should be distinguished: conceptualizing, formulating, and articulating [7, 8]. First, the speaker prepares a so-called preverbal message representation as the input for the linguistic formulation processes. It corresponds to the Text Planning of NLG. Second, as Levelt supposed, the formulation processes consist of two processing steps, grammatical encoding and phonological encoding. The grammatical encoder retrieves semantic and syntactic information and the phonological encoder retrieves the information about the words' phonological form. Its output is a phonetic plan of the utterance. The Sentence Planning seems to correspond to this processing level, especially the grammatical encoding. Third, the phonetic plan forms the input for the next processing stage, the Articulator. The Articulator executes the plan through the muscles in the speech system, resulting in overt speech, almost

the same as Realization.

There are so many works on the development and application of NLG systems [9, 10, 11]. Compared with commercial spoken dialogue systems and template-based generation system, NLG has several advantages. First, it is trainable and flexible. It based on the psycholinguistic rules and was evaluated by human judges, so it could be more and more like natural language production as the rules improved. Second, the rules for each generation module are general and domain-independent, which makes NLG be applied across domains and dialogue situations. However the quality of the output for a particular domain, or a particular situation in a dialogue, may be inferior to that of a template-based system. Last but not the least, it allows the Prosody Assignment component to have access to all of the previous levels of representation.

The most salient property of NLG systems is trainable, which gives them opportunity and room for improvement. The development of engineering technology and algorithm must speed the improving process of NLG systems. While from the psycholinguistic perspective, these rule-based systems require sophisticated linguistic knowledge. So if the achievement on cognitive researches of language production could be applied to the NLG, these systems would have a great step and be more like natural language production, as it is called. Actually, some researchers had done great job on it. In the following part, we will use an example to explain this.

2. The development of NLG

The way to develop a natural language generation system is to train its modules. Recently, several different techniques for automatically training different modules of NLG systems have been proposed [12, 13]. These techniques are mostly based on corpus analysis. While Walker, Rambow, and Rogati proposed an automatically trained sentence planner called SPoT [14], basing on feedback provided by two human judges. This methodology is unique in neither depending on hand-crafted rules nor on the existence of a domain-specific corpus.

In SPoT, sentence planning is a set of inter-related but distinct tasks, one of which is sentence scoping. Actually sentence scoping is an essential question in sentence production researches and many studies investigated it in recent decades. More and more evidence indicated that speakers have initiative in scoping the sentence and the planning scope is flexible [15]. The SPoT is trained just basing on the flexibility of sentence scoping by humans. In the next part, we will introduce the cognitive researches on sentence scoping and talk about this aspect in SPoT.

2.1. Researches on sentence scoping

The sentence scoping came originally from the incremental processing hypothesis. Kempen and Hoenkamp used an incremental procedural grammar to explain why speakers are fluent in natural dialogue with rare pauses, repetitions, and repairs, even though the processes of producing an utterance are complex. Incremental processing means that speakers can implement some plans while they are in the process of formulating others [16]. Suppose we characterize a complete sentence as a series of units. It is true that speakers must plan unit x before they articulate it. But, they can plan unit x+1 while they are articulating unit x. This makes it not necessary for speakers to complete planning the whole sentence before articulating. Otherwise, speakers will have a long pause before almost every sentence during the speech.

As a result of incrementality, one of the central issues in speech production concerns the planning units. In order to address this issue one can explore how much information the speaker plans in advance at each level before articulation is triggered. Since in NLG systems the sentence scoping only concerns the grammatical level, here we do not pay much attention on the phonological planning scope.

The grammatical planning scope was investigated by several experimental methods but still in debate. Using the picture naming task, many studies proposed a phrasal planning scope [17, 18, 19]. Meyer used picture-word interference paradigm but found a clausal planning scope [20]. Meanwhile, some eye tracking studies supposed that the grammatical encoding is processed word by word, which means that the planning scope is only the first word of the sentences [21, 22, 23]. According to the inconsistent results mentioned above, the grammatical planning scope seems to be flexible and adjusted by speakers. This was also implied in Ferreira and Swets' study [24]. They found that speakers tended to be more incremental if a deadline was set for responding, that means, the reaction after the deadline would be recorded as wrong. It revealed that the extent to which people speak incrementally is under strategic control.

2.2. Sentence scoping in SPoT

Walker et al.'s research was exactly based on the assumption that the grammatical planning scope is flexible, and trainable natural language generation is needed to support more flexible and customized dialogues with human users [14]. SPoT first randomly generates a candidate set of sentence plans and then

selects one (more details on algorithm seen in [14]). They asked human judges to compare SPoT's output with a hand-crafted template-based generation component, two rule-based sentence planners, and two baseline sentence planners, and found that SPoT performs better than the other systems. The superiority of SPoT is surely not just due to the basic assumption on sentence planning scope, but its contribution should not be ignored.

3. Outlook for the future NLG systems

In the past years, training the sentence planner of NLG system always concerned the syntactic selection and assumed that the syntactic structure is settled before retrieving every word in a sentence. Of course this processing order makes computer easy to execute and produce utterances, but the human processing is not as simple as this. For example, Bock and Levelt proposed the lexically driven models in which the thematic marking of lexical concepts affects the grammatical role marking of lemmas [25]. It means that the word retrieval could affect the syntactic selection. Moreover, several studies focused on the horizontal information flow during sentence production, and found that the information of the latter words can influence the selection of the former one [26, 27]. So the interaction between the modules of sentence planner and realizer should be concerned in the future NLG systems.

References

[1] B. W. Hardy and D. A. Scheufele. Examining differential gains from Internet use: Comparing the moderating role of talk and online interactions. *Journal of Communication*, 2006, (55), pp. 71-84.

[2] Zhu Y.. *Experimental Psychology*, Peking University Press, Beijing, 2009.

[3] R. Pieraccini and J. Huerta. Where do we go from here? Research and commercial spoken dialog systems. *6th SIGdial workshop on Discourse and Dialog,* Lisbon, Portugal, 2005.

[4] K. V. Deemter, E. Krahmer, and M. Theune. Real versus Template-Based Natural Language Generation: A False Opposition? *Computational Linguistics,* 2005, (31), pp. 15-24.

[5] E. Reiter and R. Dale. Building applied natural language generation systems. *Natural Language Engineering*, 1997, (3), pp. 57-87.

[6] S. Busemann and H. Horacek. A flexible shallow approach to text generation. In *Proc. 9th International Workshop on Natural Language Generation*, Canada, 1998, pp. 238-247.

[7] G. S. Dell. A spreading activation theory of retrieval in sentence production. *Psychological Review*, 1986, (93), pp. 283-321.

[8] W. J. M. Levelt. *Speaking: From intention to articulation.* MIT Press, Cambridge, MA, 1989.

[9] R. Dale, C. Mellish, and M. Zock. *Current Research in Natural Language Generation,* Academic Press, London, 1990.

[10] L. Cahill, C. Doran, R. Evans, C. Mellish, D. Paiva, M. Reape, and D. Scott. In search of a reference architecture for NLG systems. *Proceedings of the 7th European Workshop on Natural Language Generation,* 1999, pp. 77-85.

[11] A. Ratnaparkhi. Trainable methods for surface natural language generation. *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference,* 2000, pp. 194-201.

[12] I. Langkilde and K. Knight. Generation that exploits corpus-based statistical knowledge. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, 1998, (1), pp. 704-710.

[13] S. Bangalore and O. Rambow. Exploiting a probabilistic hierarchical model for generation. *Proceedings of the 18th conference on Computational linguistics*, 2000, (1), pp. 42-48.

[14] M. A. Walker, O. C. Rambow, and M. Rogati. Training a sentence planner for spoken dialogue using boosting. *Computer Speech and Language,* 2002, (16), 409-433.

[15] V. Wagner, J. D. Jescheniak, and H. Schriefers. On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2010, (36), pp. 423-440.

[16] G. Kempen and E. Hoenkamp. An incremental procedural grammar for sentence formulation. *Cognitive Science*, 1987, (11), pp. 201-258.

[17] W. J. M. Levelt and B. Maassen. Lexical search and order of mention in sentence production. In W. Klein & W. J. M. Levelt (Eds.), *Crossing the linguistic boundaries* (pp. 221-252), Dordrecht, Reidel, 1981.

[18] M. Smith and L. Wheeldon. High level processing scope in spoken sentence production. *Cognition*, 1999, (73), pp. 205-246.

[19] P. H. Allum and L. R. Wheeldon. Planning scope in spoken sentence production: the role of grammatical units. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 2007, (33), pp. 791-810.

[20] A. S. Meyer. Lexical access in phrase and sentence production. *Journal of Memory and Language,* 1996, (35), pp. 477-496.

[21] A. S. Meyer, A. M. Sleiderink, and W. J. M. Levelt. Viewing and naming objects: eye movements during noun phrase production. *Cognition*, 1998, (66), pp. B25-B33.

[22] A. S. Meyer and F. F. van der Meulen. Phonological priming effects on speech onset latencies and viewing times in object naming. *Psychological Bulletin and Review*, 2000, (7), pp. 314-319.

[23] Z. M. Griffin. Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 2001, (82), pp. B1-B14.

[24] F. Ferreira and B. Swets. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 2002, (46), pp. 57-84.

[25] J. K. Bock and W. J. M. Levelt. Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945-984), Academic Press,

New York, 1994.
[26] M. Smith and L. Wheeldon. Horizontal information flow in spoken sentence production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 2004, (30), pp. 675-686.
[27] J. C. Yang and Y. F. Yang. Horizontal flow of semantic and phonological information in Chinese spoken sentence production. *Language and Speech,* 2008, (51), pp. 267-284.