**A Basic Guide On Natural Language Generation:**

**S. Yashwanth1, Sneha.R2, Dinesh.T3 and Mrs. D. Agalya4**

*1UG student, B Sc. Computer Science, Department of Computer Science,*

*Jeppiaar College of Arts and Science, Padur, Chennai, India*

*E-mail:* [*s.yashwanth0428@gmail.com*](mailto:s.yashwanth0428@gmail.com)

*2UG student, B Sc. Computer Science, Department of Computer Science,*

*Jeppiaar College of Arts and Science, Padur, Chennai, India*

*E-mail:* [*rameshpushpa3209@gmail.com*](mailto:rameshpushpa3209@gmail.com)

*3UG student, B Sc. Computer Science, Department of Computer Science,*

*Jeppiaar College of Arts and Science, Padur, Chennai, India*

*E-mail:* [*tdineshtdinesh95@gmail.com*](mailto:tdineshtdinesh95@gmail.com)

*4Assistant Professor, Department of Computer Science,*

*Jeppiaar College of Arts and Science, Padur, Chennai, India*

*E-mail:* [*agalya888@gmail.com*](mailto:agalya888@gmail.com)

**Abstract:**

Applications with Natural Language Generation (i.e., like Voice Assistants) have become popular nowadays, Natural Language Generation (NLG) is defined as the process of transforming and generating written or spoken narratives from a non-linguistic input to make it easier for humans to understand, they are making it possible with the help of Artificial Intelligence programming. There are many literature reports that confessed about NLG, this is a small literature covering the basics of NLG. This literature includes procedures, tools and applications of Natural Language Generation (NLG). This research paper will help to know about the process of Natural Language Generation. Then it would also useful to researchers and students to formulate their research by knowing more about Natural Language Generation.

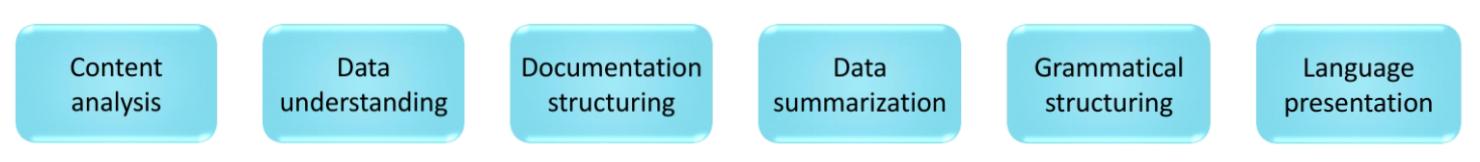
**Introduction:**

Natural Language Generation (NLG) is defined as the process of transforming and generating written or spoken narratives from a non-linguistic input with the help of artificial intelligence to make it easier for humans to understand. The process of NLG is to prepare (verbalize) some action, that the input may be non-linguistic sound or content (sound or text from a language or situation) and transform it into some linguistic form that is easier for the user to understand (written or spoken narratives). The conversion from non-linguistic data into an understandable linguistic form involves some decision-making process.

**Procedure:**

The Natural Language Generation is the multi-stage process of converting input data into output text. This process contains of six sub-process they are,

* Content analysis
* Data understanding
* Documentation structuring
* Data summarization
* Grammatical structuring
* Language presentation

It is like a pipeline process, where the output of the previous sub-process is taken as input to the current sub-process.

Content analysis:

The process of content analysis is to decide that which information should be and should not be included in the construction of output text. There is much more information in data than we want to convey it through text. Content determination chooses essential data, the choice may depend upon the users and the basic contents to be needed for the output.

for an instance,

In a cricket match we don’t want to verbalize each movement of every fielder in the ground, instead the scores, wickets and target are the essential information to be displayed.

Data understanding:

After filtering the essential information from the data, the algorithm moves on to the second step of the process – Data understanding. It is refers to the recognition of content and relate the information pattern with the datasets which are used to analyse the input. The assembling of data in a piece-by-piece manner makes the structure more efficient.

Documentation structuring:

Documentation structuring is an essential stage of NLG. Trained algorithms are used to structure the document, it is planned to bind the data together which helps in the generation of natural language easily and efficiently.

Data summarization:

Like any other stages, Data summarization is very important and is responsible for the data to be presented in a best way, it is the process of selecting relevant parts of the data that had been analysed previously and summarize the data by placing it in a good structure. The sentences are joined together and the relation between the sub-topics are initiated for the flow of the content.

Grammatical structuring:

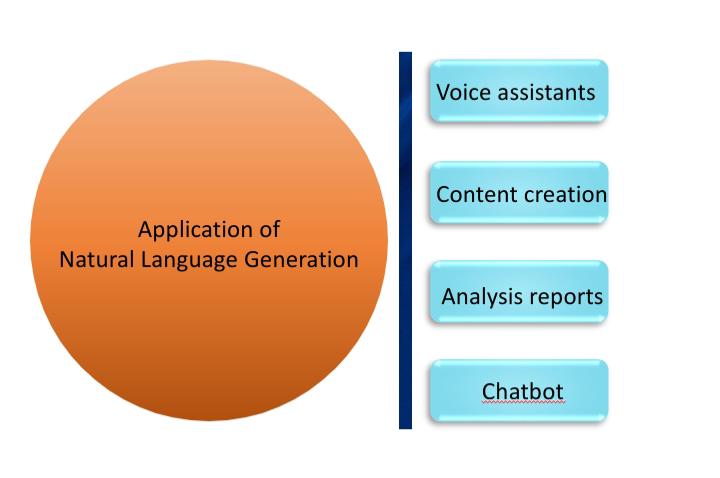
For generating a text output, grammatical construction is also a very important process as it provides a good readability to all. For an instance, Imagine reading a document with a lot of grammatical error (without appropriate conjunctions, without any punctuations, etc), which will definitely lead us to miscommunication. This stage is used to frame the grammatical structure of the output in a readable way.

Language presentation:

Language output is the last stage of the NLG, which produces output from filtered data from

the entire process. The output is produced in a structured document format that is easy for humans to read and understand anytime.

**Application of NLG:**

Nowadays The Natural Language Generation are mainly used in application and software to make the work more easier. It provides user with the information they need, which makes them understand the task to be performed.

Voice assistant:

Nowadays voice assistant are used in various fields as well as personal uses. Voice assistant use the process of natural language processing to analyse text or sound and responds to it in a written text format and also spoken narratives. Voice assistants can order our food for us, this saving us time to do other works.

Content creation:

Content creation is playing an important role in e-commerce. Many companies use content creation software that uses the NLG process, which converts statistical output into text format as it is easy for the user to improve their business. It is also used to review some products based on the features of a good product. It is also used in some paraphrasing bots.

Analysis reports:

Another eminent uses of Natural Language Generation is to generate performance analysis reports. Creating a report about a process is sill a difficult work for humans. By using NLG the computed data is converted into readable text in the basis of creating detailed report. It develops a well organized reports for an individual or an organisation or a business without consuming more time.

Chatbot:

Applications of NLG is eminently known for the chatbots communication, is the process of helping to machines to communicate with human. It also helps in the communication with customers in e-commerce platform, it converts computational data to human natural language which allow us to rectify the problem in the service.

**Tools available:**

The NLG tools or software are used to implement the process of Natural Language Generation. There are many online tools and software are available now.

* Pencil – It uses NLG to generate advertisements automatically.
* Phrasee – It uses NLG to write subject lines to the E-mail.
* Quill – It uses NLG to rephrase the content.
* Arria NLG - It uses NLG to produce written text.

These are some of the tools available in online.

**Conclusion:**

The Natural Language Generation is like a bridge that connecting human and machines, by helping the machines to generate a linguistic output which is easy for humans to understand. NLG is becoming one of the important aspects of human-machine communication and helps people in business, jobs, e-commerce communication etc.

**Reference:**

[1] John A. Bateman, Michael zock, Natural Language Generation, ResearchGate publication, January-2012.

[2] Albert Gatt, Emiel Krahmer, Survey of the State of the Art in Natural Language Generation: Core tasks, application and evaluation, Journal of Artificial Intelligence Research – 2018.

[3] Ehud Reiter, Building Natural Language Generation Systems, May-1996.

[4] Natural Language Generation Model For Mammography Reports Simulation, IEEE Journal of Biomedical and Health Informatics, September – 2020.