

# Natural Language Generation In Artificial Intelligence

Ilyas Khan

Department Of Linguistics  
Aligarh Muslim University Aligarh, India

---

## Abstract

In this article, I will explain natural language generation in artificial intelligence. Generation functionality is also being asked to play a more significant role in established applications such as artificial intelligence and machine translation. The article includes a discussion what is natural language generation and how natural language generation techniques should be used; a description of the basic NLG tasks of content determination, determining text structure and content, planning of discourse, sentence aggregation, lexicalization, referring expression generation and linguistic realization. Throughout, the emphasis is on the description of NLG that can be used to construct simple but practical working systems in artificial intelligence.

**Keywords;** natural language generation, artificial intelligence, machine translation, lexicalization,

---

Date of Submission: 27-02-2024

Date of Acceptance: 07-03-2024

---

## I. Introduction

Natural language generation in artificial intelligence is a computational approach, which reduce human efforts, meets costumer demands and find different innovative ways to keep up. Natural language generation is a subfield of computational linguistics. It focuses on computer systems that can produce understandable texts in human languages. NLG system use the application domain and the knowledge of languages to automatically produce documents, reports, explanations, help messages, and other kinds of texts.

Several companies have been started since 2009 which build systems that transform data into narrative using NLG and AI techniques. These include Arria NLG, Automated Insights, Narrative Science, Wordsmith, and Yseop, which is to be elaborated in this article.

The process to generate text can be as simple as keeping a list of canned text that is copied and pasted, possibly linked with some glue text. The results may be satisfactory in simple domains such as horoscope machines or generators of personalised business letters. However, a sophisticated NLG system needs to include stages of planning and merging of information to enable the generation of text that looks natural and does not become repetitive. The typical stages of natural language generation, as proposed by Dale and Reiter are:

For example, using the historical data for 1-July-2005, the software produces Grass pollen levels for Friday have increased from the moderate to high levels of yesterday with values of around 6 to 7 across most parts of the country. However, in Northern areas, pollen levels will be moderate with values of 4. In contrast, the actual forecast (written by a human meteorologist) from this data was.

Pollen counts are expected to remain high at level 6 over most of Scotland, and even level 7 in the south east. The only relief is in the Northern Isles and far northeast of mainland Scotland with medium levels of pollen count.

Comparing these two illustrates some of the choices that NLG systems must make; these are discussed below.

**Content determination:** Deciding what information to mention in the text. For instance, in the pollen example above, deciding whether to explicitly mention that pollen level is 7 in the south east.

**Document structuring:** Overall organisation of the information to convey. For example, deciding to describe the areas with high pollen levels first, instead of the areas with low pollen levels.

**Aggregation:** Merging of similar sentences to improve readability and naturalness. For instance, merging the two sentences *Grass pollen levels for Friday have increased from the moderate to high levels of yesterday* and *Grass pollen levels will be around 6 to 7 across most parts of the country* into the single sentence *Grass pollen levels for Friday have increased from the moderate to high levels of yesterday with values of around 6 to 7 across most parts of the country.*

**Lexical choice:** Putting words to the concepts. For example, deciding whether *medium* or *moderate* should be used when describing a pollen level of 4.

**Referring expression generation:** Creating referring expressions that identify objects and regions. For example, deciding to use *in the Northern Isles and far northeast of mainland Scotland* to refer to a certain region in Scotland. This task also includes making decisions about pronouns and other types of anaphora.

**Realisation:** Creating the actual text, which should be correct according to the rules of syntax, morphology, and orthography. For example, using *will be* for the future tense of *to be*. The Pollen Forecast for Scotland system[2] is a simple example of a simple NLG system that could essentially be a template. This system takes as input six numbers, which give predicted pollen levels in different parts of Scotland. From these numbers, the system generates a short textual summary of pollen levels as its output.

Natural Language Generation (NLG) is a subfield of artificial intelligence (AI), which produces language as output on the basis of data input, is not a new concept. What is new, however, is the increase in **adoption of NLG into the enterprise**. There are a plethora of ways the technology is being employed, primarily to improve human productivity, customer engagement and operational efficiency. People have always communicated ideas from data. But with the explosion of data that needs to be analyzed and interpreted, coupled with increasing pressures to reduce costs and meet customer demands, the enterprise must find innovative ways to keep up.

As it turns out, a machine can communicate ideas from data at extraordinary scale and accuracy. And it can do it in a particularly articulate manner. When a machine automates the more routine analysis and communication tasks, productivity increases and employees can focus on more high-value activities.

As stated in the book, "Practical Artificial Intelligence for Dummies":

*"The goal of natural language generation (NLG) systems is to figure out how to best communicate what a system knows. The trick is figuring out exactly what the system is to say and how it should say it. Unlike NLU (Natural Language Understanding), NLG systems start with a well-controlled and unambiguous picture of the world rather than arbitrary pieces of text. Simple NLG systems can take the ideas they are given and transform them into language. This is what Siri and her sisters use to produce limited responses. The simple mapping of ideas to sentences is adequate for these environments."*

Gartner's recent Hype Cycle for BI and Analytics sums up the difference between NLG and NLP (Natural Language Processing) well:

*"Whereas NLP is focused on deriving analytic insights from textual data, NLG is used to synthesize textual content by combining analytic output with contextualized narratives."*

#### **In addition to NLG types of NLP-**

**Speech-to-text:** Natural language understanding (NLU), Software which extracts information from written text, such as some of the tools that make up IBM Watson.

**Text mining, text analytics:** Natural language generation (NLG), Software which produces narratives and reports, in easy-to-read language, such as Arria's NLG Platform.

**Data in, language out:** Speech synthesis, Software which speaks or reads out text, such as CereProc CereVoice.

**Text-to-speech-** NLP applications can combine different aspects of NLP. For example, speech-to-speech applications, such as Apple's Siri and other intelligent personal assistants, use all the above types of NLP.

NLP (and hence NLG) is a type of Artificial Intelligence (AI). AI encompasses software which does "intelligent" things. What counts as "intelligent" changes overtime. Current examples of AI include (in addition to NLP and NLG): Robotics and computer vision: e.g. face recognition software and self-driving cars.

Game playing: e.g. AI players and characters in computer games.

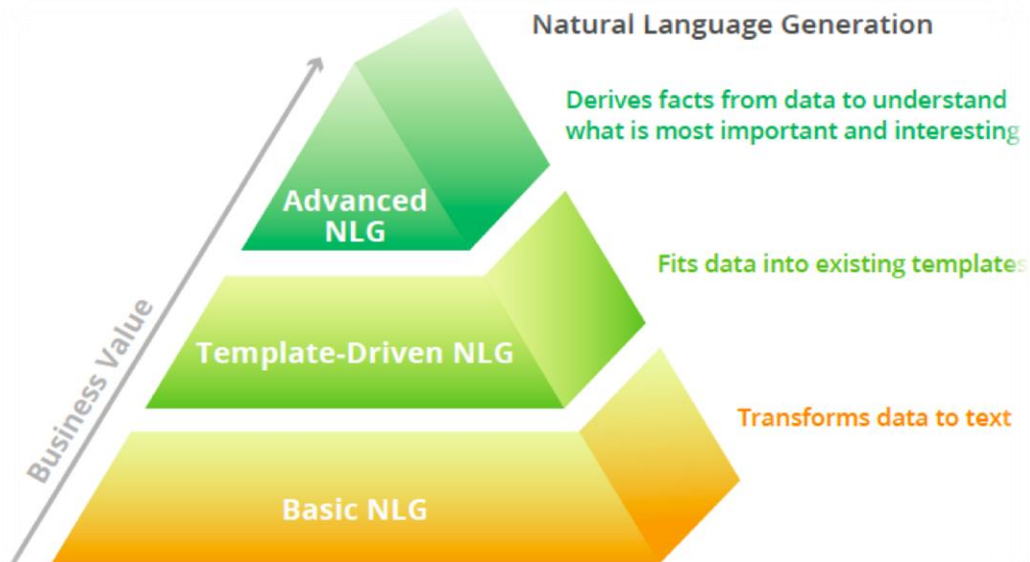
Reasoning and advisory systems: e.g. systems which recommend approving or denying mortgages, and systems which identify potential fraud in financial transactions.

In other words, **NLP reads while NLG writes**. NLP systems look at language and figure out what ideas are being communicated. NLG systems start with a set of ideas locked in data and turn them into language that, in turn, communicates them.

#### **Variations of NLG**

- **Basic NLG:** Basic NLG automatically translates data into text via Excel-like functions. An example of this would be a mail merge that restates numbers into language.
- **Templated NLG:** Here, the user is responsible for writing templates, determining how to join ideas and interpreting the output. Essentially sentence building, it relies on business rules, basic calculations (ex: sum) and templates with boilerplate text to automate content. Templated systems are limited in their ability to draw from multiple data sources, perform advanced analytics, achieve reusability from one project to the next and explain how it came to the story it created, with no understanding of what the user is trying to communicate or their particular domain.

- **Advanced NLG:** Advanced NLG communicates the way humans do – infusing intelligence and intent into the process from the very beginning. It assesses the data to identify what is important and interesting to a specific audience, and then automatically transforms those insights into Intelligent Narratives – insightful communications packed with audience-relevant information, written in conversational language. Backed by a knowledge base, Advanced NLG systems understand the domain and can write contextually about a user’s business at a scale that is not possible by humans.



NLG tools automatically analyze data, interpret it, identify the most significant parts, and generate written reports in plain English. In essence, NLG brings artificial intelligence to business intelligence (BI), automating routine analysis, saving business users time and money.

Although BI products generate visualizations, reports, and dashboards, business users still have to analyze and interpret data. That’s where NLG comes in. It automatically performs the analysis and generates an English language translation of what is significant and meaningful in the data. Business users no longer have to study the data to interpret its meaning; NLG tools do that for them.

Moreover, NLG tools bypass the need to create visualizations, charts, and reports in the first place. The tools can sift through large volumes of data and generate reports automatically. This is particularly valuable in the age of big data where huge amounts of data can overwhelm business users and IT departments alike. With NLG tools, data analysts can spend 80% of their time analyzing data rather than 80% of their time preparing data. In other words, NLG tools augment the job of business users so they can focus more on high-value tasks and less on menial work.

For example, Dominion Dealer Solutions use a NLG tool from [Narrative Science](#) to generate unique vehicle descriptions based on data gathered from a variety of sources, including automotive reviews, Kelley Blue Book, and, CARFAX. A study revealed that cars with NLG-generated descriptions sell an average of twenty days earlier than those without one.

### Crisis for human jobs

Since NLG tools automate analysis and have human-like capabilities, some people worry that NLG and other artificial intelligence products will undermine job security. NLG tools allow business users to focus on higher value tasks. Although NLG tools generate written analysis, business users still have to read, understand, and act on the reports. Rather than replacing analysts, these tools allow for more analysis to happen more often.

However, this explanation is incomplete. What if an NLG tool automates reports that usually take a dozen people a month to write? In this case, the people who are paid to prepare those reports will need to be shifted to another role in the business, if one is available, or laid off. For example, Forbes uses Narrative Science as a cost-effective way to generate content and expand market coverage. Instead of hiring people, Forbes chose Narrative Science to generate the content.

## **Here is an overview of three major NLG products today:**

### **Narrative Science**

Based in Chicago, Narrative Science sells Quill, an NLG product favored by financial services companies, although the product can be used by companies in any industry. “If there is structured data and there needs to be something communicated from that data, it’s a use case for Quill”, says Mary Grace Glascott, Director of Product Marketing at Narrative Science.

For example, a large financial services company uses Quill to provide automated feedback to call center representatives. Quill analyzes call detail data and automatically generates a personalized performance report written in English for thousands of customer service representatives, something that is difficult to do with traditional BI tools.

Initially, Quill was only available in the cloud, requiring users to send data out and receive analysis back. However, the product now also runs on-premises. Narrative Science also sells extensions that enable Quill to run within major BI platforms, including Qlik Sense, Tableau, SAP, Business Objects, Lumura, and Microsoft Power BI.

### **Yseop’s Savvy**

Based in Dallas, Texas, Yseop sells Savvy, an NLG tool which plugs into any data warehouse or database. Savvy specializes in helping companies leverage their CRM data to generate reports and narratives that make it easy for sales people to cross-sell and upsell products.

When using Savvy with a CRM application, such as [Salesforce.com](https://www.salesforce.com), business users select the account they want to analyze, and then click on an Yseop tab to see a written report. Matt Rauscher, vice president of Yseop, says, “Savvy takes data from a CRM application and its rules engine automatically decides, based on the data, what products a salesperson should sell to which customers, and then the NLG tool writes what they need to do and why”.

Unlike other tools, Savvy can currently write in multiple languages, which is a requirement for global enterprise companies. Also, Yseop positions Savvy as a development platform for building NLG-based applications.

### **Arria’s NLG Platform and Recount**

Arria offers NLG Platform, which works with almost any data source or application. NLG Platform runs on the cloud or on-premises and works in any industry, including financial services, healthcare, and marketing. Soon, Arria will release two cloud versions of the product: an enterprise version called Articulator Pro and Articulator Lite geared to non-NLG programmers that will compete with Yseop Savvy.

Arria also sells Recount, an accounting solution that works with accounting software from Xero. Recount automatically generates reports that identify key trends and issues in their accounting data that they need to pay attention to. The tool helps small and medium-sized business owners focus on the business rather than administrative duties and bookkeeping.

According to Jeff Zie, CMXO and Head of Recount, “Xero is a fantastic accounting platform, but it doesn’t tell owners what’s going on and whether it’s a good thing or a bad thing. Using Recount is like picking up a phone to a financial analyst and asking ‘what’s going on with my business?’”

## **So, What’s The Takeaway?**

NLG tools automate analysis, taking the capabilities of BI tools to the next level. Rather than generate charts and tables, NLG tools interpret the data and generate analysis in written form that tells business users what’s significant to know. The tools perform routine analysis of predefined data sets, eliminating both the manual labour required to generate reports and the skilled labour required to analyze and interpret the results.

Although NLG vendors say their tools augment, not replace, jobs of report writers and analysts, in some cases, NLG tools will reduce the number of people required to generate and analyze data. For business owners, this is great news: they can reassign staff to other jobs, increase their productivity, or lay them off.

The main requirement for implementing NLG is the ownership and access to a structured dataset. In order for any natural language generation software to produce human-ready prose, the format of the content must be outlined and then fed structured data.

For example, our Wordsmith platform accepts structured data via uploading a CSV directly, passing the data to our API as a JSON object, or through one of our integrations that connect to the API like Zapier or Tableau.

## **NLG Use in Technology**

While NLG can be implemented wherever there is a need to generate content from data, some of the most common uses of the technology include: generating product descriptions from inventory data, creating

individual financial portfolio summaries and updates at scale, business intelligence performance dashboard text explanations, real estate property descriptions and personalized customer communications.

Every NLG narrative is designed to read as though a human wrote it. The specific data insights and writing style of each narrative vary widely depending on the specific case and audience. The structure and tone of each piece of content is determined by the person designing the narrative structure, or template.

This is why NLG is flexible enough to generate data-driven narratives about industries ranging from finance and business intelligence to e-commerce and sports.

### **Advantages of Using NLG**

**Save Time and Money:** With NLG, you can produce thousands of unique narratives in the time it takes to write a single piece of content. For example, the process of writing hundreds or thousands of product descriptions is time-consuming and costly—whether that's stealing valuable time from in-house writers or hiring an army of freelancers to complete the task.

### **Wordsmith Product Description**

With NLG, each row in a structured dataset, like a CSV of inventory specs, can be transformed into a unique, variable description. This process can occur at scale, generating thousands of unique, variable descriptions for every item in an e-commerce store's inventory.

### **Extensive Personalization at Scale**

NLG enables you to generate complex personalization at scale, creating improved customer communication and experience with your organization.

A financial services firm, for example, can deliver individualized portfolio summaries to thousands of clients while reducing the cost of production.

Natural language generation works for companies with both a large established customer base and for companies expecting rapid growth of their client pool. That's because NLG enables businesses to make full use of existing data while allowing for the addition of new client data to produce unique content for each individual customer, without an exponential increase in cost. Check out our Orlando Magic case study on improved customer communications through personalization.

### **Make Data Understandable and Useful**

Charts and graphs can be visually overwhelming and often require additional explanation from the experts manipulating the supporting data. Natural language conveys expert-level analysis and advice in concise, plain terms that engage and fully inform each reader. For example, NLG enables every sales person on a large team to have a performance report tailored precisely to them, benchmarking their individual performance against their peers and giving suggestions for improvement.

NLG can also be integrated into existing business intelligence dashboards so that the generated text changes in tandem to different visualization filters.

### **Tableau visualization with Wordsmith NLG explanation**

Our Wordsmith platform, for example, is integrated with Tableau so that users can also generate custom, responsive narrative explanations of their data visualizations. The narrative can be customized to use a business's unique terminology, as well as edited to highlight the most important metrics for each of the company's departments. Check out live, responsive examples of our Tableau integration.

### **NLG In The Future**

In the past, natural language generation implementations would take months of work by professional data scientists and software developers. A business might invest thousands of dollars in NLG solutions without seeing any content for several quarters.

Today, our Wordsmith platform makes it easy for anyone to upload data and start automating their own consumer-facing stories in a few hours. Wordsmith users can also take advantage of many integrations to quickly implement NLG in their existing processes. And don't be fooled by the nomenclature—"template" NLG is just as complex as "advanced" natural language generation. The content produced by both methods will be as intricate as the writer constructing the software logic makes it.

Alexa, Cortana and others are ushering in the era of intelligent personal assistants, helping to make everyday tasks easy and efficient for consumers. The enterprise is catching up, with **conversational interfaces** that are facilitating engagement across employees and to customers, raising the bar on how these systems communicate.

What is the critical differentiator from a conversational bot that performs tasks to one that engages, explains, and illuminates? Advanced NLG.

Per a recent article in the Harvard Business Review, "Bots that Can Talk Will Help Us Get More Value from Analytics":

*"Conversations with systems that have access to data about our world will allow us to understand the status of our jobs, our businesses, our health, our homes, our families, our devices, and our neighbourhoods — all through the power of Advanced NLG. It will be the difference between getting a report and having a conversation. The information is the same but the interaction will be more natural."*

Like any subsection of computer software, the functionalities of what software vendors call "NLG software" varies, with some even trying to commercialize simple templating systems which generate narrative like "mad libs" (yes, I know I am dating myself with that reference).

Basically, these simple NLG systems have prewritten sentences with holes in them. For example, "The weather in <<location>> is <<temp>> degrees". These simple systems turn one data point into a sentence which works for some very use cases but aren't scalable, easily maintainable or generally enterprise ready.

NLG software is used to automate the writing of written reports, but more broadly it fits into the data to data-driven decision-making workflow. NLG is really the last mile in this process. Data collection and analysis have already generally been automated; it's the last step of explaining the results of the analysis in plain English (of whatever language) where NLG helps.

## II. Conclusion

Today, Wordsmith platform makes it easy for anyone to upload data and start automating their own consumer-facing stories in a few hours. Next-generation NLG is software able to summarize larger amounts of data and able to explain WHY numbers are what they are. These systems allow companies to generate both back-office and front-office reports since they write more than just descriptive narrative. In fact, they go further with tools that are able to explain analysis, like Yseop Compose, which writes in multiple languages and installs on a private cloud or on premise.

## References

- [1]. "Factset Uses Precisely Structured Data And Yseop Software To Write Corporate Descriptions". Www.Forbes.Com. Retrieved 2017-01-31.
- [2]. Dale, Robert & Reiter, Ehud (2000). Building Natural Language Generation Systems. Cambridge, U.K.: Cambridge University Press.
- [3]. Evans, Roger; Piwek, Paul; Cahill, Lynne (2002). "What Is Nlg?". New York, Us.
- [4]. Goldberg E, Driedger N, & Kittredge, R (1994). "Using Natural-Language Processing To Produce Weather Forecasts". Ieee Expert. 9 (2): 45–53. Doi:10.1109/64.294135.
- [5]. Harris, Md. (2008). "Building A Large-Scale Commercial Nlg System For An Emr". Proceedings Of The Fifth International Natural Language Generation Conference. Pp. 157–60.
- [6]. Perera R, & Nand P. (2017). "Recent Advances In Natural Language Generation: A Survey And Classification Of The Empirical Literature". Computing And Informatics. 36 (1): 1–32.
- [7]. S Sripada, N Burnett, R Turner, J Mastin, D Evans(2014). Generating A Case Study: Nlg Meeting Weather Industry Demand For Quality And Quantity Of Textual Weather Forecasts. Proceedings Of Inlg 2014
- [8]. Sauper, Christina & Barzilay, Regina (2009). "Automatically Generating Wikipedia Articles: A Structure-Aware Approach". Proceedings Of Acl.
- [9]. Yuko Sakai (2017). Sentence Generation: Syntax Tree Diagram In English, Spanish, Chinese, Japanese, Ainu. Isbn 978-1545429006