

# A Commercial Perspective on Reference

**Ehud Reiter**  
Arria NLG PLC  
ehud.reiter@arria.com

## Abstract

I briefly describe some of the commercial work which Arria NLG is doing in referring expression algorithms, and highlight differences between what is commercially important (at least to Arria) and the NLG research literature. Arria's focus is on high-quality algorithms for types of reference which are important in its systems. These algorithms need to be parametrizable for different genres and domains, usable in hybrid systems which include some canned text, and support variation.

## 1 Introduction

There is an extensive academic literature in NLG on generating referring expressions. In this paper I partially describe the types of reference which are important to Arria NLG, a company which builds commercial NLG systems (I cannot fully describe what Arria does because of commercial confidentiality).

In general terms, the high-level concepts behind Arria's work are similar to the high-level concepts behind academic NLG work. However there is a difference in emphasis, and hence in the specifics of algorithms. In particular, Arria has focused less on the task of identifying salient visual or physical entities, and more on specialized reference tasks such as referring to a specific component in a complex machine, and referring to a company in a contextually appropriate way. Arria also wants its reference algorithms (and indeed all of its NLG algorithms) to support a number of practical criteria:

- *Configuration*: Easily configurable and parametrizable for different genres and domains.
- *Hybrid NLG/template systems*: Usable in systems which produce documents which include canned text as well as NLG text.
- *Variation*: Allow random or systematic variation (when desirable), so users who regularly read generated texts don't see the same referring expression used again and again.

In this paper I will give some examples of the reference algorithms Arria has developed, and explain how they meet the above criteria.

## 2 Background: Reference

Referring expression generation has been a focus of NLG research since the 1990s (van Deemter 2016a); a good recent survey is Krahmer and van Deemter (2012). Much of this research has been on choosing definite NPs (such as "the dog" or "the big black dog") to refer to physical objects which are already salient to the hearer. As described by Krahmer and van Deemter, many algorithms have been developed for this task, and there also has been work on data sets and evaluation criteria, and a shared task (e.g., Gatt and Belz 2008). A substantial amount of work has also been done on using pronouns, and on generating references to sets. Less research has been done on reference to non-physical entities such as dates or companies. In terms of the above criteria:

- *Configurability*: Some algorithms are parametrisable, for example the incremental algorithm (Dale and Reiter 1995) allows a genre/domain specific preference order to be specified between features. But this has not been a focus of research.
- *Hybrid systems*: Similarly some work has been done on reference in systems which include canned text (e.g., van Deemter et al 2005, Belz and Kow 2010), but this has not been a research focus.
- *Variation*: This has been addressed indirectly via research (motivated by cognitive modelling) on probabilistic reference algorithms (e.g., Gatt et al 2013, Mitchell et al 2013).

In short, while the criteria of interest to Arria have been addressed in academic research, they have been peripheral and not the main focus of this work.

### 3 Background: Arria

Arria NLG is a company which specializes in selling NLG solutions and technology, especially data-to-text systems. As described on Arria's webpage<sup>1</sup>, Arria uses a fairly standard data-to-text NLG pipeline (Reiter 2007). This pipeline is incorporated into Articulator Pro (A-Pro), which is Arria's NLG software development kit (SDK). One of Arria's systems was described (including evaluation) in an earlier INLG paper (Sripada et al, 2014).

Most of Arria's systems generate texts which are intended to support professionals such as engineers, doctors, and financial analysts. Thus, Arria focuses on language used in professional contexts, not everyday language.

### 4 Reference at Arria

A-Pro has a generic API for reference modules. This means that different reference modules can be plugged into a system, depending on what is being referred to (e.g., person, place, time, company, machine, etc.), and the genre. Reference modules can access a domain model (which describes reference

targets) and a discourse model (which records linguistic context). Below I briefly describe some of the specific reference modules which Arria has developed for A-Pro.

It is of course essential that Arria's reference algorithms be fast computationally, robustly implemented and tested, well documented, and interface easily to external data sources and domain models. I will not further discuss such software engineering issues in this paper, but they are very important.

#### 4.1 Component Reference

Arria has developed and indeed obtained a patent on a reference algorithm for components in complex machinery (Reiter, 2016). This algorithm arose out of work that Arria did in the oil industry, where it was necessary to refer to specific components in a complex machine in a narrative text which described the status of the machine. The specific context is confidential and also quite complex, but a related problem is referring to body parts (Fig 1). For example, suppose a mother is talking to her three children, Ann, Bob, and Charlotte, and wishes to refer to the index finger of Ann's left hand. Depending on the discourse context and previous utterances, the mother could say

1. The index finger of Ann's left hand
2. The left hand index finger
3. The index finger
4. It

For instance, reference (1) would be appropriate in a null context, or if the previous utterance had been about Bob. Reference (2) might be appropriate if the previous utterance had been about Ann's face. Reference (3) would make sense if the previous utterance had been about Ann's left hand, and reference (4) could be used if the previous utterance had been about the Ann's left-hand index finger.

Arria's algorithm assumes there is domain model which specifies a part-of hierarchy of the machine (or body) in question, and a discourse model which keeps track of previous references to entities in the domain model. When a new reference is needed, the algorithm essentially looks for the lowest common parent of the most recent previous referent and the new reference target, and constructs a referring expression by traversing the part-of hierarchy from the common parent to the reference target.

<sup>1</sup> www.arria.com

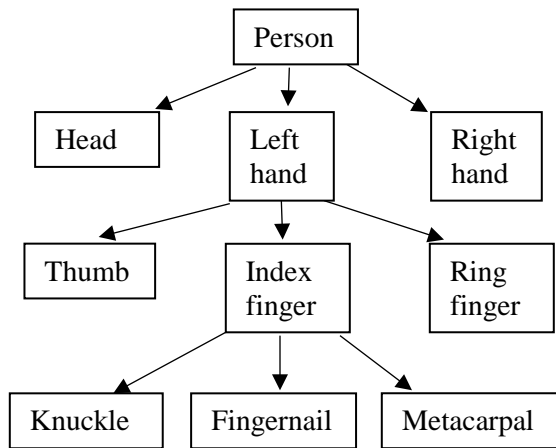


Fig 1: part-of hierarchy for body (extract)

For example, if the previous reference was (1) below, then the algorithm might produce (2)

(1) Ann’s left thumb is scratched.

(2) The index finger is bleeding.

In this case

*PreviousRef*: thumb of left hand of Ann

*TargetRef*: index finger of left hand of Ann

*Lowest common parent*: left hand of Ann

*PartofHierarchy from parent to referent*: index finger

*Referring expression*: index finger

In terms of the criteria mentioned above

- *Configurability*: At the semantic/content level, the algorithm allows levels in the part-of hierarchy to be skipped, and special names to be used. For example, we can configure the algorithm so that the thumb of the left hand is referred to as the “left thumb”, not the “thumb of the left hand”. Realisation of referring expressions (e.g., the maximum number of noun-noun modifiers) can also be configured.
- *Hybrid systems*: Excluding pronouns, the algorithm works as long as all component references are generated via the algorithm; everything else can be canned text. In other words, the algorithm can be used with structures such as “I am worried about [X]”,

where X is a component reference and everything else is canned text.

- *Variation*: This is supported by allowing algorithm to occasionally start from a higher node than the lowest-common parent (e.g., produce “left-hand index finger” instead of “index finger”, even if the latter is sufficient in the context), and to vary realization (e.g., “the index finger of the left hand”).

## 4.2 Named Entity Reference

Arria has also developed an algorithm for referring to named entities such as companies. This is very important in financial services, which is one of the sectors which Arria is targeting.

For example, suppose that a financial report wished to refer to Arria as a company. Should it say

1. Arria NLG
2. Arria
3. It

Reference (1) would be appropriate when the company was first mentioned in a text, or when the full name was contextually required. Reference (2) would be appropriate when the company had already been introduced in the text, and a short name was unambiguous. References (3) would be appropriate when the discourse context made it clear what the pronoun referred to. Note that the algorithm needs access to an external data source of name variants, otherwise it would not know, for example, that International Business Machines and IBM referred to the same entity.

The algorithm basically looks for the shortest referring expression which works in the current discourse context. Crucially, it is customizable for different genres and clients. For example, some genres require a full legal name (e.g., Arria NLG PLC), and in other genres a stock name (e.g. GOOGL) should be used to refer to a company.

Appropriate use of pronouns also depends on genre and client. In particular, some clients are relatively “relaxed” about pronoun usage, because they think semantic context will disambiguate pronoun references; however for other clients pronouns should only be used if there is no possibility of confusion. For example, consider “Yahoo had a poor year. It may need a new CEO”. Using “It” to refer

to Yahoo is acceptable under a relaxed strategy which assumes that semantic context will rule out “a poor year” as a potential reference target. However under a strict reference policy “it” could not be used here, since (at least from a purely syntactic perspective) it could refer to the year.

From the perspective of the above criteria

- *Configurability*: supporting configurability (including pronoun strategies) is the most complex aspect of the algorithm.
- *Hybrid systems*: Similar to the previous algorithm, template structures such as “I recommend buying [X]” can be used provided that all company name references are generated via the algorithm.
- *Variation*: The algorithm can be configured so that a specific form cannot be repeated more than N times in a row.

### 4.3 Time and date reference

Arria also has an algorithm for time and date reference; date reference in particular is very important in financial reporting. This algorithm allows timestamps to be referred to at different levels of granularity (e.g., minute, day, year), using discourse-appropriate references. For example, if granularity is day, then the timestamp 00:00:00 28 April 2017 could be referred to as

1. 28 April 2017
2. 28 April
3. the next day

Formatting can be configured, for example we can get April 28, 2017 in USA. In any case, reference (1) could be used in a null context, reference (2) in a context where the previous date reference was to another day in 2017, and reference (3) when the previous date mentioned in the text was 27 April 2017.

From the perspective of the above criteria

- *Configurability*: Developers can control which forms are allowed in the text (which depends on genre), as well as formatting.
- *Hybrid systems*: Similar to the previous algorithms, templates such as “I went to New York on [X]” can be used provided that all time/date references are generated via the algorithm.

- *Variation*: The algorithm can be configured to vary the forms used in a specific context.

## 5 Discussion

High-quality referring expressions are important to Arria, in part because they distinguish Arria’s systems from text-generation systems built with non-NLG technology. However from Arria’s perspective, academic research on generating referring expressions has been less useful than originally anticipated. What would be ideal from Arria’s perspective is research on specific types of reference which are common in the domains Arria works in, focusing on algorithms which are sensitive to linguistic and discourse context, configurable, usable in hybrid systems which include some canned text, and which support variation.

There are definitely encouraging signs, for example the recent resurgence of interest in contextually appropriate named entity reference (e.g., Belz and Kow 2010, van Deemter 2016b), although this has mostly focused on people rather than companies. It is also encouraging to see recent work on variation (e.g., Baltaretu and Ferreira 2016) and on configuring reference for different genres and domains (e.g. Koolen et al, 2012).

Of course NLG researchers do not need to focus on Arria’s needs. But there are many interesting research issues in specific types of reference, variation, etc. Also human speakers arguably use different reference strategies for different types of entities, vary reference strategies depending on domain and genre, insert referring expressions into fixed (formulaic) language, and vary reference in order to keep text interesting. Investigating these issues could lead to important insights about language and reference.

### Acknowledgements

Many thanks to the many people on the Arria team (too many to list here) who have worked on developing, testing, and documenting the above-mentioned algorithms. My thanks also to Kees van Deemter and other members of the Aberdeen University CLAN research group for their very valuable advice and suggestions; and to the anonymous reviewers for their comments and suggestions.

## References

- A Baltaretu and T Ferreira (2016). Task demands and individual variation in referring expressions. In *Proceedings of INLG-2016*.
- A Belz and E Kow (2010). The GREC Challenges 2010: Overview and Evaluation Result. In *Proceedings of INLG-2010*.
- R Dale and E Reiter (1995). Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions. *Cognitive Science* 19:233-263
- K van Deemter, E Krahmer, M Theune (2005). Real vs. template-based natural language generation: a false opposition? *Computational Linguistics* 31:15-24.
- K van Deemter (2016a). *Computational Models of Referring*. MIT Press.
- K van Deemter (2016b). Designing Algorithms for Referring with Proper Names. In *Proceedings of INLG-2016*.
- A Gatt and A Belz (2008). Attribute Selection for Referring Expression Generation: New Algorithms and Evaluation Methods. *Proceedings of INLG-2008*.
- A Gatt, R van Gompel, K van Deemter, E Kramer (2013). Are we Bayesian referring expression generators. *Proceedings of CogSci 2013*.
- E Krahmer and K van Deemter. 2012 Computational Generation of Referring Expressions: A Survey. *Computational Linguistics* 38:173-218
- R Koolen, E Krahmer, M Theune (2012). Learning Preferences for Referring Expression Generation: Effects of Domain, Language and Algorithm. *Proceedings of INLG-2012*.
- M Mitchell, K van Deemter, E Reiter (2013). Generating Expressions that Refer to Visible Objects. *Proceedings of NAACL-2013*, pages 1174-1184.
- R de Oliveira, S Sripada, E Reiter (2015). Designing an Algorithm for Generating Named Spatial References. *Proceedings of ENLG-2015*, pages 127-135
- E Reiter (2007). An Architecture for Data-to-Text Systems. *Proceedings of ENLG-2007*, pages 97-104.
- E Reiter (2016). Method and apparatus for referring expression generation. US Patent 9355093.
- S Sripada, N Burnett, R Turner, J Mastin, D Evans (2014). A Case Study: NLG meeting Weather Industry. *Proceedings of INLG 2014*.
- R Turner, S Sripada, and E Reiter (2010) Generating Approximate Geographic Descriptions. In E Krahmer and M Theune (eds). *Empirical Methods in Natural Language Generation*, pages 121-140. Springer.