

Identifiable Indefinites and the Givenness Hierarchy: a Case of Underspecification and Overspecification?

Katy Jones (jonesks1@cardiff.ac.uk)

School of English, Communication and Philosophy
Cardiff University, Colum Drive, Cardiff, CF10 3EU. UK

Abstract

This paper analyses a particular type of indefinite nominal phrase within the framework of the Givenness Hierarchy (GH). It tests the GH's claims that referring items can attain a higher cognitive status than their coding signals, and are merely underspecified for higher statuses. Results from two reader interpretation experiments largely support the GH analysis; readers mostly interpret these structurally indefinite expressions as referring to a specific, fully-identified entity. I then discuss reasons for the underspecification and consider whether, in cases when the conceptual data in the expressions is very specific, the expression could also be lexically overspecified, as the content is more detailed than is necessary for unique identification.

Keywords: Identifiable indefinites; the Givenness Hierarchy; underspecification; overspecification; interpretation of referring expressions.

Introduction

Conventionally, the definite article is recognised as being used to refer to entities that are 'given', while the indefinite article is for entities which are not identifiable to the addressee. However, some indefinite expressions appear not to follow this principle. In [1], it is possible to understand '*a precious green field that...*' as a newly introduced referent. But once it is seen in context, where 18 mentions of this particular field (Rumney Recreation Ground) have already been made, it becomes clear that the indefinite article is signalling something different.

[1] [...] a precious green field that has been the "green lungs" of life in east Cardiff for almost a century

The purpose of this paper is twofold. The first aim is to analyse expressions such as [1] within the implicational framework of the Givenness Hierarchy (Gundel, Hedberg and Zacharski 1993 – GHZ forthwith). It tests the claims made by the Givenness Hierarchy that referring expressions can attain a higher cognitive status than their encoding signals, and may be just underspecified for higher statuses. Further, the underspecification may be a result of the lexical content of the expression or of its intended interpretation within its specific context. Through an empirical analysis of 40 opinion articles from British broadsheet newspapers using the coding protocol for statuses on the Givenness Hierarchy (GHZ 2006), I test whether these expressions meet the criteria for cognitive statuses which are higher than their coding conventionally indicates. I show that, in certain contexts (e.g. in British English journalistic opinion writing), an expression of the type A(n)+NOUN + RESTRICTIVE RELATIVE CLAUSE (A(n)+N+RRC hereafter), such as [1], attains a higher cognitive status in the Givenness Hierarchy than it conventionally signals.

The second aim of this paper is to provide empirical evidence from two reader interpretation experiments. The purpose of these experiments was to determine whether readers, when interpreting the cognitively underspecified A(n)+N+RRC expression, have access to the representation of the entity that they have built up through the previous references to it, or whether they create a new 'type' representation that draws on procedural information encoded in the indefinite expression. The results suggest that there is a 'scale of specificity' with regard to the A(n)+ N+RRC expressions, depending on the amount and detail of known conceptual information in the restrictive relative clause.

I then consider whether the more lexically specific instances of the A(n)+N+RRC expression, such as [1], exemplify overspecification as well as underspecification, as the conceptual information in the relative clause is more detailed than is necessary for unique identification. Further, overspecification may facilitate identification and account for the high number of readers interpreting the more specific expressions as referring to identifiable entity.

Data

As stated above, most theories of (in)definiteness agree that definiteness involves some sort of shared knowledge between the speaker and the addressee and a referent is considered definite if the addressee is able to identify it (e.g. Hawkins 1978; Chafe 1994; Givón 1993; Radden & Dirven 2007). On the other hand, an indefinite expression is considered to be one whose referent the addressee is not familiar with or is unable to identify.

However, consider [2] from an opinion article about footballer David Beckham, in which all the definite references to Beckham are italicised:

[2] *David Beckham*: how *this crock of a footballer* can still woo the French.

It doesn't matter *he's* too old to play for Paris Saint-Germain – *Becks* can still score in France as an honorary Frenchman [...]. It's the kind of astute diplomatic move that, along with *his pretty face, good manners and chiselled abs*, makes one suspect *Beckham* isn't British at all. It seems that France, like Spain and the United States before it, is poised to be charmed by *a man who, with his un-British attention to grooming, muscle tone and non-novelty underwear, may become an honorary Frenchman before his six months in Paris are up*.

These expressions are unproblematic as they are clearly definite in their form and function. However, a problem arises when considering the final nominal phrase in this paragraph (in bold and italics). Structurally, this expression is indefinite as it contains the indefinite article *a*, but when considered in its context, there is noticeably something different at play. Firstly, consider the textual and discourse

environment: references to David Beckham have been made 18 times previously, two of which are in the headline, and two in the sub-head. Also, there is an image of David Beckham below the head and sub-head. Further, after the appearance of this expression, there are an additional 14 definite references to Beckham. By the time the indefinite expression occurs, the reader has built up a solid discourse representation of Beckham, a representation which also continues to be developed afterwards.

Now consider the conceptual content of the expression. The information in the relative clause is textually given: the reader has been informed of Beckham's penchant for taking pride in his appearance and the fact that he may become an honorary Frenchman previously in the text. The textually new piece of information, about the *non-novelty underwear*, is homophorically given. That is, anyone interested enough in Beckham to be reading this article will also know that he has a line in designer underwear which he models himself. Thus, it would be difficult to *not* connect the semantic content of the relative clause to the definite referent.

So there is something of a paradox here: there is a formally indefinite expression, which is traditionally viewed as being non-identifiable, but which is part of an identity chain of references to a definite referent, whose semantic content is given and is so specific that it can only really be referring to one particular entity; that is, the previously mentioned definite referent. The conditions for the use of the definite article are met, but the indefinite article is chosen instead. So how does the reader know how to interpret the expression?

The data in this investigation comprises 40 opinion articles about a specific entity (e.g. a famous person or well-known organisation) from British broadsheet newspapers, within which 45 instances of the expression type A(n)+N+RRC occur. As in the Beckham example, the data in the RRCs is textually given, and in the few instances where it is not, it is homophorically given. These expressions differ from the late indefinites examined by Epstein (1994: 219-226) as the content of the RRCs in his data is new, and so the identifiable referent is 'reclassified' in light of the new information attributed in the relative clause. For example, consider [3], taken from Epstein (1994: 224):

[3] [...] an end to the trade embargo that has shredded the economy of a nation that already was the poorest in the western hemisphere

Epstein comments that the referent 'nation' has been put in a category to which it has not previously been portrayed as belonging, hence the 'newness' (c.f. Schouten & Vonk 1995; Ushie 1986). The indefinite expressions in this study do not have new information in the RRCs, but given and thus they cannot be explained though a 'reclassification' of an existing referent.

So the question is how is an analyst to approach this kind of expression, given that most theories of (in)definiteness would explain it as a previously unmentioned denotatum.

Accessing memory representations of referents

Perhaps one way to approach these kinds of identifiable indefinites is to consider them in terms of what they are doing cognitively or procedurally; that is, the accessing of referents in terms of memory representations.

For Ariel (1990), this is in terms of accessibility; the speaker's job when referring is to select a referring expression by taking into account the assumed level of accessibility of the mental entity for the addressee. However, the focus is solely on definite descriptions and indefinite expressions do not feature in the Accessibility Marking Scale. Givón also puts forward a gradation of referring expressions which reflects that of Ariel, and does mention indefinites. However, 'referential indefinite NPs' (1983: 17) are considered to be 'maximally surprising/disruptive/discontinuous' (1983: 25). That is, referential indefinite NPs are thought to be for entities which are newly introduced into the discourse, and not for subsequent mentions. Thus, in both these theories, the A(n)+N+RRC expression here would have to be dealt with as a marked form for marked use, which is reasonable but beyond the scope of this paper.

The Givenness Hierarchy

A cognitive account of referring which may be able to explain this particular use of an indefinite expression for definite referent without assuming it is a marked form for marked use is the Givenness Hierarchy. The Givenness Hierarchy (GHZ 1993) is a framework in which lexical items in referring expressions conventionally signal both procedural information about the assumed cognitive status of the referent in the mind of the addressee and conceptual information. That is, the intended interpretation of a referring expression is determined not only by procedural information about location and manner of interpretation but also by the descriptive or conceptual information encoded in the expression (GHZ 1993; 2012).

The Givenness Hierarchy outlines six implicationaly-related cognitive statuses which are relevant to the form of referring expressions. The hypothesised English forms and the statuses they are assumed to encode are outlined below. The most specified forms are the highest forms on the left.

Table 1: The Givenness Hierarchy

in focus	activated	familiar	uniquely identifiable	referential	type identifiable
>	>	>	>	>	
{it}	{that, this this N}	{that N}	{the N}	{indefinite this N}	{a N}

As the GH is an implicational scale, each status entails the statuses to its right (i.e. lower statuses). Thus, if an expression encodes the status 'familiar', it also entails that the expression is 'uniquely identifiable', 'referential' and 'type identifiable'. Due to lack of space, a detailed explanation of the different statuses is not given here, but see GHZ (1993:276-280) for a full account.

As we have seen, the expressions under review here are formally indefinite, and thus would typically fall under the status of 'type-identifiable'. That is, the addressee is expected to be able to access a type representation of the entity conveyed in the referring expression, but not the identity of the specific thing (GHZ 1993: 276). So in the following example [4] from the current data, the addressee should be able to identify the type of institution that has such attributes:

[4] *An institution that allows the maintenance of a stained glass ceiling for its female clergy to bang their heads against* should not only lose its moral authority; let it also lose its unearned privileges.

However, the argument here is that the addressee should be able to do more than access a type representation; the particular institution has had several previous mentions and the information in the relative clause makes the referent uniquely identifiable in the given context. The GH seems to provide an explanation for this: as an implicational scale, the statuses on the scale are not mutually exclusive (GHZ 2012). In using a particular form, the speaker not only indicates that the relevant cognitive status has been met, but that all lower statuses have been too. Thus in [4], the referent may simply attain the cognitive status 'type-identifiable' as its coding would typically signal, or it could be 'referential', 'familiar', 'uniquely identifiable', 'in focus' or 'activated' as these statuses all entail 'type-identifiable'. Therefore, forms which appear to encode a particular status are simply underspecified for higher statuses, and are not excluded from them (GHZ 2012: 5).

One possible explanation for this use of the indefinite article when a higher status is met is that given the detailed descriptive content of these expressions, information about a higher status is unnecessary for identifying the intended referent (GHZ 2012: 6). In cases like [5], using a full nominal expression rather than a pronoun (which would have been appropriate as the relevant cognitive status is attained) permits extra descriptive information to be added to the expression.

[5] *His* is a fall from grace like no other. *A man who can dismiss the US anti-doping agency's finding of "conclusive and undeniable proof" of cheating as a tax-payer funded hatchet job* probably doesn't do remorse.

Using a pronoun or a full nominal (Lance Armstrong) here would result in the loss of the descriptive content of the expression and thus the informative value would be changed. And perhaps it is the informative value that is most important here, as it may be necessary for pragmatic or rhetorical purposes.

So, cognitive status is not the sole determinant for the use of an appropriate referring expression; there are a number of interacting factors which contribute to choice of expression type, both linguistic and extralinguistic. Thus, even though use of a particular determiner or pronoun implies that a stronger form does not hold in a certain context, it does not exclude a higher cognitive status; the forms are simply underspecified for them.

The GH Coding Protocol

If, as predicted by GHZ, indefinite expressions are simply underspecified for cognitive statuses higher than 'type-identifiable', it seemed appropriate to carry out an analysis within the framework of the GH to determine what cognitive statuses these particular expressions actually attain. To do this, the Coding Protocol for Statuses on the Givenness Hierarchy (GHZ 2006) was employed. The GH Coding Protocol is a list of criteria for each cognitive status on the hierarchy. In order to establish which status applies, the analyst simply puts herself in the position of the speaker/writer and considers what she can assume about the

addressee's cognitive status of the intended referent at the point before the occurrence of the form in question. The analyst then reads through the list of criteria and stops once she has reached a criterion which applies to the particular expression in question; that is the highest cognitive status which that particular referring expression can attain (GHZ 2006: 1).

Due to space constraints, the full list of criteria is not reproduced here. However, the criteria which apply to the current data can be found below (for a full list of the criteria, see GHZ 2006). Examples from my own analysis have been used to illustrate each criterion.

In Focus:

1. It is the interpretation of the main clause subject or the syntactic topic in the immediately preceding sentence/clause (see [5] above).
2. It is part of the interpretation of a previous part of the same sentence.

[6]From *his complete letters* a remorseless self-portrait emerges of *a man who from his youth onwards found it very difficult to coexist with others or find a productive place in his 19th-century world*.

Activated:

3. It is part of the interpretation of one of the immediately preceding two sentences.

[7] You can find it in the Regulators' Compliance Code for *Natural England*. It is not clear to me why *an agency whose stated aim is to defend the environment* should have to "encourage economic progress" [...]

Familiar:

4. It was mentioned at any time previously in the discourse. (This criterion applies to all the data).

As all of the expressions in this data had been mentioned several times previously in the discourse before the occurrence of the A(n)+N+RRC expression, in theory there was no need to look any further than the criterion for 'familiar' (4). However, in order to ensure that the GH is not obscuring other possibilities by the way it distributes the options, I read through all the criteria for lower statuses to confirm other options do not apply more appropriately. Interestingly, as predicted by the GH, at least one criterion for each status lower than 'familiar' can be applied easily to all the expressions. The first criterion for 'uniquely identifiable' is pertinent here ('the referring form contains adequate descriptive/conceptual content to create a unique referent') as I am arguing here that the lexical content of the expression guides the reader's interpretation to an identifiable referent.

Following the instructions laid out in the GH Coding Protocol, I read through the criteria for each status until a criterion which applied to the particular expression applied. If there was some uncertainty as to whether a particular criterion applied, caution was exercised and the next criterion was considered instead, finally settling on one which could be more confidently applied. Admittedly, the judgements made by the analyst are subjective, and before any firm conclusions can be drawn from this investigation,

the coding should be carried out by at least one other coder. That said, it cannot be disputed that the expressions here attain at the very least the status of ‘familiar’ as it is unquestionable that the referent has been mentioned previously in the discourse, so the argument that these expressions are underspecified remains strong. The only issue in question is which status above ‘familiar’ they attain.

A summary of the results is laid out below. (A full table of the complete results is available on request. This includes the title of each text, the expression within its context, the cognitive status the particular expressions attained as well as the specific criterion that applied to it):

Table 2. Coding results

Cognitive Status	# of expressions (/45)
Familiar	15
Activated	10
In focus	20

As predicted, the A(n)+N+RRC expressions are underspecified for the statuses they attain, according to an analysis using the GH coding protocol. They are, at the very least, familiar (33%) due to their having been mentioned in the discourse previously. The remaining expressions either attained the status of activated (22%) or in focus (44%). There were no instances of expressions only attaining the status type-identifiable, as the form conventionally signals. So it seems that the GH can provide a description for the use of these expressions; their cognitive status is underspecified. Possible reasons for this are discussed in the final section.

Although the GH can provide an explanation for the use of these expressions, it would be interesting to clarify this by determining whether readers’ interpretations correspond to the GH analysis. So the next stage of this investigation was to set up an experiment in which readers had to determine whether the (A(n)+N+RRC) expression referred to a ‘type’ or the specific, identifiable entity (which would assume a cognitive status of ‘uniquely identifiable’ or above).

Reader interpretation experiments

The purpose of the initial experiment was to determine whether readers, when interpreting the underspecified A(n)+N+RRC expression, access an existing representation of the identified entity or whether they create a new representation based solely on the conceptual information in the expression. Due to space constraints, the experiments are described very briefly here.

Experiment 1 Hypothesis: Readers interpret the A(n)+N+RRC expression as referring to the specific, identifiable referent.

The participants were made up of 91 2nd year undergraduate Language and Communication students and were divided (almost) equally into three groups. Each group saw three texts (about Van Gogh, Nadine Dorries and Rumney Rec) and each text occurred in its original or a modified form, so that it reflected, in different versions, definite, indefinite and type. No group saw the same text twice or the same condition twice. For example, Group 1 saw Text 1 in the indefinite A(n)+N+RRC condition (which reflected the original text), Text 2 in the forced ‘type’ condition PLURAL N+RRC and Text 3 in the forced definite condition This+N+RRC. Effort was made to ensure

that the modified texts remained as close to the original texts as possible. Groups 2 and 3 saw the same texts but in different conditions. Participants were asked to decide which of the multiple choice answers they thought applied to the underlined expression. [8] is an example of the tasks for Group 1:

[8]

Text 1(indefinite ‘a’ condition)

The MP Nadine Dorries has proposed an amendment to a health bill demanding that women be given pre-abortion counselling that could be provided by religious, pro-life groups. Dorries is basing her bill on the premise that teaching abstinence stops teenagers becoming pregnant. The fact that an MP who can spread such inaccuracies is proposing a nonsensical, victim-blaming bill should give us all cause for concern.

- The type of MP who can broadcast untruths
- The MP Nadine Dorries
- Any MP
- None of the above

Text 2(‘type’ condition)

Rumney Recreation Ground is a historic green space at the heart of the Rumney area, but it is under threat because the council wants to build a new high school there. People tend to understand that there is a need for the development and construction of new 21st century schools. But these schools don't have to be built on precious green fields that have been the "green lungs" of life in cities for a long time.

- The type of land which provides urban green space for city dwellers
- Any precious green field
- Rumney Recreation Ground
- None of the above

Text 3(‘definite’ condition)

The new Penguin edition of Van Gogh’s complete letters is dispiriting because it reveals the seriousness and extremity of Van Gogh’s emotional troubles. From his complete letters a remorseless self-portrait emerges of this man who from his youth onwards found it very difficult to coexist with others or find a productive place in his 19th century world.

- Van Gogh
- Any man
- The type of man who was not comfortable with his 19th century world
- None of the above

Results:As it is the indefinite ‘a’ condition we are interested in, the responses for the other conditions are not discussed here (however, it can be noted that almost all participants chose the ‘definite’ answer in the ‘definite’ condition and the ‘type’ answer in the ‘type’ condition). In the indefinite condition, 67 participants selected the specific entity (74%), 23 the type answer (25%) and one chose the ‘any’ option (1%). Chi-Square was applied: chi= 111.8; p< 0.001.

Thus, the results support the hypothesis that readers generally interpret the A(n)+N+RRC expression as referring to the specific entity rather than a ‘type’, therefore supporting the GH analysis that these expressions attain a higher cognitive status than their coding signals. However, what is interesting about these results is that there was not an equal distribution of ‘type’ answers among the three texts

(i.e. 7 or 8 for each text), as might be expected if all the A(n)+N+RRC expressions were structurally and conceptually similar. Instead, the ‘Nadine Dorries’ text triggered two thirds of the ‘type’ responses (15/23). On closer examination, it became obvious that the conceptual data in the relative clause of this expression is less detailed and explicit than the other two texts. Compare:

- [9] an MP who can spread such inaccuracies
- [10] a precious green field that has been the “green lungs” of life in east Cardiff for almost a century
- [11] a man who from his youth onwards found it very difficult to coexist with others or find a productive place in his 19th-century world

Clearly, the relative clauses in [10] and [11] contain considerably more ‘definiteness’; that is, their content can be easily attributed to a specific referent, whereas the content of [9] is more general and could apply to a type of MP. So it seemed appropriate to set up another experiment to test whether there may be a ‘scale of specificity’ with regard to the A(n)+N+RRC expressions.

Experiment 2 Hypothesis: The texts with very specific semantic information in the RRC are more likely to be interpreted as referring to the specific entity in the target ‘a’ condition, whereas those with less specific semantic content are more likely to be interpreted as referring to a ‘type’ in the ‘a’ condition, irrespective of the presence of the indefinite article.

Participants in experiment 2 comprised six groups of 31st year undergraduates (n=186). There were six texts: three with very specific and detailed relative clauses (about Louis Kahn, Andy Coulson and David Beckham), and three with more general information in the relative clauses (about the Jewish National Fund, Natural England and Gaelic). Experiment 2 was otherwise set up in exactly the same way as experiment 1; participants read three texts in three different conditions, and the texts, conditions and multiple choice answers were all counter-balanced.

Results: Again, only the results for the indefinite condition are discussed here. As predicted, the three texts which contained the specific RRCs generated a high number of specific responses (75/93 - 81%; Chi=160.032, p < 0.001), whereas those with the more general RRCs produced an approximately equal distribution of answers between type and specific (46/93 - 49% and 47/93 - 51% respectively; Chi= 74.957, p < 0.001).

So the results suggest that there is a ‘scale of specificity’ with regard to the A(n)+ N+RRC expressions, depending on the amount and detail of given conceptual information in the restrictive relative clause. The more specific the details in the relative clause, (i.e. the more the information can be tied uniquely to the participant the text is about), the more likely it is that the reader interprets the indefinite noun phrase as referring to the given, identifiable entity. On the other hand, when the details in the relative clause are less explicit readers are equally divided between interpreting the expression as referring to the specific entity and a ‘type’, in the latter case creating a new ‘type’ representation.

Discussion and conclusions

The results of the reading experiments largely support the GH analysis. The expressions with the detailed lexical

content are interpreted as referring to an identifiable referent, which would correspond to the statuses of ‘uniquely identifiable’ and above (remember though that according to the GH protocol, these expressions attain at the very least the higher status of ‘familiar’ due to their having previously been mentioned in the discourse). However, those with semantically less explicit information in the RRCs appear to create more ambiguity; around half of the participants interpreted them as referring to a ‘type’ (which corresponds to the status of ‘type-identifiable’), and half as referring to the identifiable entity (‘uniquely identifiable’ and above).

The results of the reading experiments suggest that it is perhaps the conceptual content encoded in the relative clause within these expressions which guides the addressee to their interpretation rather than the presence or absence of a definite determiner. But when an expression contains insufficient conceptual information for unique identification, the reader relies on both the procedural information encoded in the item and the conceptual content for its interpretation; hence the mixed results for these less specific expressions.

Another question that needs to be addressed is why a writer might select expressions that underspecify the cognitive status in this way. As mentioned earlier, underspecification allows the writer to add more descriptive information about the object being referred to. In all the expressions here, the indefinite noun is post-modified with a restrictive relative clause, and each differs in the amount of detail about the intended referent. It is possible that this detail makes it unnecessary to provide procedural information to the addressee about a higher cognitive status as it would be redundant in terms of identification of the referent.

Consider again the expression below (reproduced from example [2]):

- [12] [...] *a man who, with his un-British attention to grooming, muscle tone and non-novelty underwear, may become an honorary Frenchman before his six months in Paris are up.*

The properties attributed to the referent ‘a man’ are so detailed and specific that they cannot apply to any referent other than the one the text is about, David Beckham. This is supported by the fact that the information in the RRC is given information; that is, it has been mentioned previously in the text and the writer just seems to be reminding the reader of these properties. It is also true that one of the conventional functions of the indefinite article is to attribute qualities to a referent (c.f. Givón 1993: 101; Burton-Roberts 1976: 428), and so the use of the article ‘a’ here, while underspecifying cognitive status, allows the writer to do this.

Further, in several of the instances of A(n)+N+RRC, the immediate cotext renders it virtually impossible for the expression to refer to anyone/thing else than the referent the text is about. Consider [13]:

- [13] From *his complete letters* a remorseless *self-portrait* emerges of a man who from his youth onwards found it very difficult to coexist with others or find a productive place in his 19th-century world.

In [13] the A(n)+N+RRC expression is the object of the nominal phrase ‘a remorseless self-portrait emerges’. The

fact that it is ‘self-portrait’ means that the ‘a man who...’ expression can only be referring to the referent of ‘a remorseless self-portrait’; the writer of ‘his complete letters’, Van Gogh. ‘Self’ cannot refer to any other person. The GH analysis supports this by showing that the nominal phrase [13] is actually ‘in focus’, as it is the interpretation of the main clause subject in the immediately preceding clause (GHZ 2006: 1).

What is also interesting with these particular expressions is that the conceptual data in the relative clauses varies, as we have seen. In some expressions, such as [14], the content of the RRCs is less specific, and could arguably be referring to a type of such-and-such.

[14] *An agency which should be protecting the natural world* appears to have identified and aligned itself with people damaging it.

If taken out-of-context, the indefinite expression in [14] could conceivably be referring to a type of agency, and the properties expressed in the relative clause could be attributed to a number of agencies which have that particular role. However, within its context, that particular agency has been mentioned 32 times previously and the GH analysis shows that it attains the cognitive status of ‘activated’, as it is part of the interpretation of one of the immediately preceding two sentences (GHZ 2006: 3). So even when the conceptual content of the relative clause is less specific, which may lead to a ‘type-identifiable’ interpretation by the addressee, information in terms of recency of previous mention signals that the expression is intended to be interpreted as belonging to a higher cognitive status.

Another curious issue here concerns the expressions which have very detailed and specific relative clauses. Consider [15]:

[15] [...] *a man who died in a public lavatory in a low-grade public building, whose corpse lay unrecognised in a New York City morgue for three days, and who flitted from one family home to another*[...]

It is possible that the expressions in this study with extremely detailed and specific relative clauses represent cases of overspecification, as well as underspecification of cognitive status. That is, the conceptual information in the relative clause appears to be more than is necessary for uniquely identifying the object the expression refers to (c.f. Arts, Maes, Noordman & Jansen 2011). As well as the expression attaining a cognitive status of ‘activated’ (as it is part of the interpretation of one of the immediately preceding two sentences), the vast detail summarising and reminding the reader of the qualities of Louis Kahn is more than necessary for the reader to uniquely identify the referent. The reader would presumably be able to do so with much less detail, particularly given the cognitive status it attains and the text upstream.

Final remarks

Irrespective of what the language conventionally encodes, in all the instances of A(n)+N+RRC, the writer creates a tension, through underspecification of cognitive status, between the choice of indefinite article and the fact that the addressee is able to identify the referent; the lexical items in these referring expressions force the reader to consider

the fully-identified referent as an unidentified instance of a collection of properties. How much this affects identification of the referent seems to depend on the amount and detail of lexical information in the relative clause. In any case, this tension seems to be deployed as a pragmatic or rhetorical device, to create a new layer of meaning that is not directly encoded in the grammar nor entailed in the cognitive status.

References

- Ariel, M. (1990). *Assessing noun-phrase antecedents*. London and New York: Routledge.
- Arts, A., Maes, A., Noordman, L. & Jansen, C. (2011). Overspecification facilitates object identification. *Journal of Pragmatics*, 43, 361-374.
- Burton-Roberts, N. (1976). On the generic indefinite article. *Language*, 52(2), 427-448.
- Chafe, W. (1994). *Discourse, Consciousness and Time*. USA: University of Chicago Press.
- Eptsein, R. (1994). *Discourse and definiteness: synchronic and diachronic perspectives*. San Diego: University of California.
- Givón, T. (1983). Topic continuity in discourse: an introduction. In Givón, T. (ed), *Topic continuity in discourse: a quantitative cross-language study*. Amsterdam and Philadelphia: John Benjamins.
- Givón, T. (1993). *English Grammar*. Amsterdam and Philadelphia: John Benjamins.
- Gundel, J.K., Hedberg, N. & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69, pp. 274-307
- Gundel, J., Hedberg, N. and Zacharski, R. (2006). *Coding protocol for statuses on the Givenness Hierarchy*. Available at: www.sfu.ca/~hedberg/coding_for_cognitive_status.pdf. Accessed 16/01/13
- Gundel, J., Hedberg, N. and Zacharski, R. (2012). Underspecification of cognitive status in reference production: some empirical predictions. *Topics in cognitive science*: 4(2): 249-68
- Hawkins, J.A. (1978). *Definiteness and indefiniteness*. London: Croom Helm Ltd.
- Radden, G., & Dirven, R. (2007). *Cognitive English Grammar*. Amsterdam/Philadelphia: John Benjamins.
- Schouten, C.H. & Vonk, W. (1995). *On the use of marked indefinite noun phrases*. Technical Report, ESPRIT Project 6665 DANDELION. Tilburg / Nijmegen etc.: DANDELION Consortium and CEC
- Ushie, Y. (1986). Corepresentation – a textual function of the indefinite expression. *Text*, 6, 427-446.