

Reference to Discourse Topics: Introducing “Global” Shell Nouns

Fabian Simonjetz
Linguistics Department
Ruhr-Universität Bochum
Bochum, Germany
fabian.simonjetz@rub.de

Abstract

Shell nouns (SNs) are abstract nouns like *fact*, *issue*, and *decision*, which are capable of referring to non-nominal antecedents, much like anaphoric pronouns. As an extension of classical anaphora resolution, the automatic detection of SNs alongside their respective antecedents has received a growing research interest in recent years but proved to be a challenging task. This paper critically examines the assumption prevalent in previous research that SNs are typically accompanied by a specific antecedent, arguing that SNs like *issue* and *decision* are frequently used to refer, not to specific antecedents, but to global discourse topics, in which case they are out of reach of previously proposed resolution strategies that are tailored to SNs with explicit antecedents. The contribution of this work is three-fold. First, the notion of global SNs is defined; second, their qualitative and quantitative impact on previous SN research is investigated; and third, implications for previous and future approaches to SN resolution are discussed.

1 Introduction

Traditionally, the primary concern of anaphora resolution (AR) has been the systematic identification of coreference between pronouns and NPs whereas less prototypical forms of anaphoric relations such as bridging anaphora and reference to abstract objects still remain comparatively unexplored (Poesio et al., 2016). One notable exception to this are *shell nouns* (SNs; Schmid, 2000), abstract nouns such as *fact*, *decision*, or *issue*, whose idiosyncratic referring properties have received growing research attention in the last couple of years.

The defining characteristic of SNs lies in their capability to refer to abstract, proposition-like entities usually expressed by non-nominal syntactic constituents such as full sentences (example 1), *that*-clauses (example 2), or infinitive clauses (example 3).¹

- (1) If the subject prefers to look at one stimulus rather than another we can assume that he has detected a difference between them. **This idea** had a great influence upon the thinking of Schopenhauer who followed up its implications more thoroughly than did Kant.
- (2) It must have seemed clear to him at once that Tolkien was a man of literary genius, and **this fact** only brought home to him his own sense of failure as a writer.
- (3) The Lake District Planning Board has sought to limit new houses to local people, but **this attempt** was overturned by the Secretary of State for the Environment.

The way SNs are interpreted in unison with syntactic clauses from their context bears a striking resemblance to anaphora, and it is this observation that sparked efforts to approach SNs from a computational perspective with the ultimate goal of automatically detecting SNs alongside their respective antecedents. This task – termed *SN resolution* (Kolhatkar, 2015) – is usually approached in the same way as traditional

This work is licensed under a Creative Commons Attribution 4.0 International License. License details: <http://creativecommons.org/licenses/by/4.0/>.

¹SNs are printed in boldface, and antecedents are underlined. The examples (1–3) are drawn from the British National Corpus (<http://www.natcorp.ox.ac.uk>).

AR, i.e., the context of a SN instance is scanned for (non-nominal) antecedent candidates which are then passed to an ML-based ranking algorithm that determines the best match. However, the wide range of syntactic shapes SN antecedents can take on adds significantly to the complexity of the problem which, along with other factors, makes major simplifications necessary so as to make the task more feasible.

Part of the difficulty of quantitative approaches to SNs has been attributed to the existence of so-called “long-distance antecedents” (Kolhatkar, 2015, p. 38). On the grounds of examples like (4–5), Kolhatkar argues that SNs can occur far away from the antecedents they refer to, and accordingly, allows her system to obtain antecedent candidates from up to four sentences preceding the SN instance plus the sentence hosting the SN.

- (4) The sense of a public struggling with a morally difficult issue was dramatically conveyed when the survey asked: “Would you approve or disapprove of someone you know having an abortion?” Thirty-nine percent said they would approve and 32 percent said they would disapprove. But 25 percent more volunteered a response not included in the question: they said their view would depend on the circumstances involved. An additional 5 percent did not know. The lack of a clear majority for either of the unequivocal responses to **this question** may be the best indicator of where public opinion really stands on abortion.
- (5) New York is one of only three states that do not allow some form of audio-visual coverage of court proceedings. Some lawmakers worry that cameras might compromise the rights of the litigants. But a 10-year experiment with courtroom cameras showed that televised access enhanced public understanding of the judicial system without harming the legal process. New York’s backwardness on **this issue** hurts public confidence in the judiciary...

Long-distance antecedents do not only make the task of SN resolution significantly more complex while offering only a rather small potential benefit, they are also problematic from a theoretical perspective because they stand in contrast to research on pronominal anaphora that found non-nominal antecedents to be of low salience, and hence, to be accessible for subsequent anaphoric reference for only a limited amount of time. Kolhatkar (2015, p. 38) attributes the seemingly increased anaphoric range of SNs to their richer semantics when compared to pronouns, arguing that the additional information conveyed by the nouns makes it possible for less recent antecedents to be identified. However, this does not seem to fully explain the remarkably long distances reported for some SN instances. For example, Kolhatkar (2015, p. 61) mentions a case where the antecedent occurs six sentences prior to the SN, and the longest distance I am aware of is an instance of the SN *issue* with an (albeit nominal) antecedent eleven sentences back in the data by Simonjetz and Roussel (2016)², which raises the question of whether such examples are adequately analyzed as anaphora.

A hint at a possible answer to this question can be found in work by Ariel (1988) who found NPs that refer to *discourse topics* (DTs), i.e., the central entities the discourse is about, to be viable antecedents for subsequent pronominal reference for longer distances than would normally be expected, indicating that DTs are highly accessible even if they have not been mentioned recently. Aside from NPs, DTs can also be expressed by questions or propositions (Watson Todd, 2016), and if it is possible for expressions as semantically unspecific as pronouns to refer to non-local antecedents as long as they correspond to DTs, it seems likely that SNs, which often carry a “topic-like” meaning (e.g., *theory*, *idea*, *issue*, *question*, and – obviously – *topic*), are suitable devices to refer to (propositional) DTs. Thus, the SNs *question* and *issue* in examples (4) and (5) can be taken to refer to the topics their respective discourses are about.

Much unlike the obligatory antecedents of SNs like in examples (1–3), DTs are accessible for anaphoric reference regardless of whether or not they are explicitly realized in the discourse. Even if an adequate description of the DT is present as is the case in examples (4–5), it appears that the marked spans of text lack the direct anaphoric relation present in examples (1–3). Accordingly, we can draw a distinction between

1. *Proper* antecedents that are locally available for immediate reference via a SN or pronoun, and

²The data is available at <https://github.com/ajroussel/shell-nouns-data> and the example can be found in turn t_02-06-11_37.

2. *Apparent* antecedents that are merely descriptions of what the SN in question refers to – often, the DT – without standing in a direct anaphoric relation.

As realizations of two distinct underlying processes, topic-referring SNs (henceforth *global SNs*) on the one hand, and SNs with antecedents in their close vicinity (*local SNs*) on the other, have different properties, which renders SN resolution strategies as previously proposed much less, if at all, adequate for the resolution of the former.

The remainder of this paper will specify the details of the implications of this observation, starting with an overview of related work in SN annotation and resolution (Section 2), followed by an examination of the (qualitative and quantitative) impact of global SNs on the tasks of SN annotation and resolution (Section 3). Next, implications for previous and future research are discussed (Section 4) and complemented by some concluding remarks (Section 5).

2 Background

2.1 Schmid’s Definition of Shell Nouns

Schmid (1997; 2000; 2018) defines SNs as a class of abstract nouns that feature an “inherent semantic gap” (Schmid, 2018, p. 111) which has to be filled with context-specific information generally represented by non-nominal syntactic units such as *that*- and *to*-clauses, or sentences and longer stretches of the discourse. The central observations underlying the notion of SNs are not new; anaphoric or deictic links between (pro-)nouns and abstract entities have been an area of interest within the linguistic and philosophical literature for more than half a century. Among the plethora of similar notions are *container nouns* (Vendler, 1967), *general nouns* (Halliday and Hasan, 1976), *situation reference* (Fraurud, 1992), *reference to abstract objects* (Asher, 1993), *discourse deixis* (Webber, 1988; Webber, 1991), *carrier nouns* (Ivanič, 1991), *labels* (Francis, 1994), *signalling nouns* (Flowerdew, 2003; Flowerdew and Forest, 2015), *abstract anaphora* (Dipper and Zinsmeister, 2012), *non-nominal antecedent anaphora* (Roussel et al., 2018; Kolhatkar et al., 2018), and others. What is innovative about SNs (and what makes them appealing for computational linguists) is their syntactically driven definition: while related notions largely rely on theoretical and philosophical considerations that are hard to operationalize, SNs are defined by Schmid (2000, p. 3) in terms of a set of syntactic patterns (examples 6–7) which serve as a fairly objective (though not perfect) linguistic test to identify SNs and as templates for corpus queries to automatically gather SNs.

- (6) Determiner + (Premodifier) + SN + postnominal *that*-clause, *wh*-clause or *to*-infinitive
The (deplorable) fact that I have no money.
- (7) Determiner + (Premodifier) + SN + *be* + complementing *that*-clause, *wh*-clause or *to*-infinitive
The (big) problem was that I had no money.

Schmid (2000) based his considerations about the theoretical and cognitive aspects of SNs on data retrieved from the Cobuild corpus by means of the patterns (6–7). The automatic resolution of SNs, however, has not been attempted until more than a decade later when SNs were recognized as a stepping stone to approach anaphora with non-nominal antecedents.

2.2 Shell Noun Annotation

Prior to implementing automatic SN resolution systems, it is necessary to manually generate gold standard data for the purposes of investigating the properties of SNs as well as establishing a data base for training and testing systems later on. Kolhatkar and Hirst (2012; 2014) and Kolhatkar (2015) explore both crowd-sourcing and expert annotations of the SN *issue* and other nouns in Medline abstracts and New York Times articles. Another expert annotation was published by Simonjetz and Roussel (2016) who present a study of parallel English and German SNs in Europarl data. Occasionally, SNs are annotated as a subset of other forms of anaphora (Poesio and Modjeska, 2002; Poesio and Artstein, 2008; Flowerdew and Forest, 2015), and there is an extensive amount of relevant literature on the annotation

and resolution of pronouns with non-nominal antecedents, overviews of which can be found, e.g., in Dipper and Zinsmeister (2010), Roussel et al. (2018), and Kolhatkar et al. (2018).

The manual annotation (and the automatic resolution, for that matter) of SNs broadly consists of three steps:

1. Identification of relevant SN instances
2. Identification of the source sentence hosting the antecedent
3. Identification of the exact antecedent within the source sentence

The first step is relatively straightforward as SNs are a semi-open class, i.e., while an exhaustive list of all SNs would be difficult to compile, there are a few hundred lexemes that are widely used as SNs (Schmid, 2000). Thus, finding relevant SNs in a corpus is for the most part a matter of matching tokens or lemmas with a list of SNs, as carried out by Kolhatkar and colleagues, who look at a number of SN lexemes that occur with an accompanying demonstrative *this* and manually exclude irrelevant instances in postprocessing (Kolhatkar and Hirst, 2012; Kolhatkar et al., 2013; Kolhatkar and Hirst, 2014; Kolhatkar, 2015). A similar approach has been pursued by Simonjetz and Roussel (2016) and Roussel (2018), i.e., an exhaustive annotation of SNs without preselecting specific nouns has not been attempted so far.

After the target nouns have been determined, annotators are instructed to identify the sentence hosting the antecedent of the SN. This step does not pose great practical difficulty either, as preprocessing essentially consists of sentence-splitting the data. Kolhatkar and Hirst (2012) and Simonjetz and Roussel (2016) do not treat the selection of an antecedent source sentence as a separate step, but Kolhatkar et al. (2013), Kolhatkar and Hirst (2014), and Kolhatkar (2015), found higher agreement for this simpler subtask than for the annotation of exact antecedent spans, indicating that annotators will often agree with respect to the approximate location of the antecedent.

The most challenging part of SN resolution is the identification of exact antecedent spans which, accordingly, received the most research attention. While antecedents in traditional anaphora and coreference resolution are for the most part restricted to NPs, SN antecedents can come in a variety of syntactic shapes, are not necessarily continuous, and may even span multiple sentences. As elaborated by Kolhatkar (2015), treating all syntactic constituents as markables is impractical as it would result in a high number of candidates most of which are either very unlikely or overlap significantly with each other, which could cause confusion among annotators. Thus, most previous work opted for allowing free spans of text to be annotated as SN antecedents, resulting in a more straightforward, yet harder to evaluate task.

Due to the lack of a shared set of items to choose from, freely annotated data cannot be evaluated on a binary hit-or-miss basis – reliability metrics need to account for and quantify the degree of overlap between annotators. In wide use are variants of Krippendorff’s (unitizing) α (Krippendorff, 2013), chance-corrected reliability coefficients specifically designed for the annotation of free spans of text that have been employed to determine the reliability of SN antecedent annotation, e.g., by Kolhatkar and Hirst (2012), Kolhatkar et al. (2013), Kolhatkar (2015), and Simonjetz and Roussel (2016). Still, the agreement estimates returned by Krippendorff’s α and other metrics of inter-annotator agreement are difficult to interpret and an imperfect solution, as they fail to incorporate semantic knowledge which would be desirable for the task of SN resolution. To illustrate, consider example (5), where two annotators could arrive at the same understanding of the SN, yet annotate distinct spans of text, e.g., “audio-visual coverage of court proceedings” versus “courtroom cameras”. This would result in a low agreement even though the selected spans of text virtually mean the same thing. Likewise, the very same amount of overlap between two non-identical annotations could either reflect a different interpretation of the SN (if the antecedents describe fundamentally different concepts) or an insignificant difference if, e.g., annotators agree with respect to the core concept but decided to include or dismiss a (possibly lengthy) adjunct.

2.3 Shell Noun Resolution

The first approach to automatically resolve the SN *issue* has been offered by Kolhatkar and Hirst (2012). Their contribution consists of an annotation of instances of *this issue* in Medline abstracts followed by

an automatic resolution approach that consists of a candidate extraction and ranking procedure. With the annotation showing reliable results, the candidate ranking model was trained and evaluated based on a range of syntactic and semantic factors as well as the distance between the SN and the candidate. This work was later expanded by Kolhatkar et al. (2013), Kolhatkar and Hirst (2014), and Kolhatkar (2015) in particular, by successively expanding the range of included SN lexemes, increasing the amount of data, and improving the ML-based resolution. On the basis of the findings and data from these studies, at least two more approaches to SN resolution have been published, namely Marasovic et al. (2017), a neural network approach which covers a wider range of SNs, and an approach to resolve German SNs by Roussel (2018).

Even though some progress has been made in the resolution of SNs, the approaches presented to date are still far from a full-fledged, general purpose SN resolution system. Domains are mostly restricted to Medline abstracts and NYT articles, and a number of additional simplifications have been adopted, e.g., by limiting the task to a few SN lexemes. In addition, different studies deal with different subtasks of the resolution. Marasovic et al. (2017), e.g., focus on the identification of the exact antecedent span within the source sentence (step 3 above) and treat the antecedent’s source sentence itself as given.

Due to the problems posed by annotating SNs, the open questions regarding best practices to deal with the inherent vagueness of SN usage, the absence of a large-scale, general purpose data set for SNs, and the lack of an agreed upon evaluation metric, a comparison of previous work on SN resolution is difficult (Kolhatkar et al., 2018). The notion of global SNs as shown below may help to partly disentangle this complex picture.

2.4 Discourse Topics

In contrast to the well researched notion of sentence topics, *discourse topics* are a concept without a generally accepted formal definition (Ariel, 1988; Asher, 2004; Watson Todd, 2016). As for many other applications, a “pretheoretical” (Watson Todd, 2016, p. 9) notion of DTs as ‘descriptions of what a discourse or discourse section is about’ will suffice for our considerations, though.

According to Watson Todd (2016, p. 50), there are three common ways to express DTs, i.e., propositions, questions, and NPs. As SNs generally refer to “proposition-like pieces of information” (Schmid, 2000, p. 4), it appears reasonable to adopt a definition that views DTs as propositions in line with, e.g., the seminal work by Asher (1993) and Asher and Lascarides (2003) in the framework of SDRT. Occasionally, it will be necessary to expand this view to questions, concepts, events, and the like, but the details of the semantic types of SN antecedents are of no concern here and have been subject to detailed examinations elsewhere (Fraurud, 1992; Asher, 1993; Schmid, 2000; Kolhatkar, 2015, and others). For instance, the topics of the discourses in examples (4) and (5) can be informally characterized as a question and as the concept of allowing trials to be filmed, respectively.

Specifying the topic(s) of a given discourse is a challenging task that defies a simple formalization (Watson Todd, 2016). The topics of the examples above are rather straightforward to identify, but this is usually not the case for more complex discourses. That being said, deciding whether or not a given SN is global or local does not require the intended topic to be specified, hence the question of how to approach DTs formally is not of primary concern for this work.

What is important for the argument put forward here is the fact that DTs, though an integral part of discourse structure, are systematically left implicit (Asher and Lascarides, 2003; Asher, 2000; Asher, 2004). This is dramatically illustrated by the Question under Discussion (QUD) view of discourse structure (Roberts, 2012) which models all utterances of a discourse as answers to underlying questions the speaker seeks to address, and while these questions may be explicitly expressed as in example (4), they usually are not. Global SNs then are SNs that refer to DTs for which the presence of an antecedent cannot be presupposed, which renders them principally out of reach of a resolution via a candidate extraction and ranking strategy. Even if an appropriate topic description is available, coincidentally matching descriptions must not be mistaken for proper antecedents as in examples (1–3) (cf. Section 4).

3 Global Shell Nouns

3.1 Global Vs. Local Shell Nouns

In Schmid's (2000) patterns, the constituent subcategorized by the SN functions as syntactically determined antecedent by conveying a proposition that occupies the argument slot inherent in the semantic structure of the SN. As illustrated by the examples (1–3), the SN can alternatively derive its argument from its nearby context, sometimes even across a sentence boundary. What is common to all these cases is that a proposition (or other abstract object) is temporarily salient enough to be available for anaphoric reference and gets immediately picked up by a SN to construct a unified meaning in a classical compositional semantic fashion. In that respect, such *local* SNs follow the same principles as pronouns that refer to non-nominal antecedents, and it is likely that they can be solved by means of similar strategies by identifying the antecedents that are accessible when the anaphor is expressed and, subsequently, selecting the best match(es).

The apparent antecedents of *global* SNs, on the other hand, can occur several sentences away from the SN, contradicting the well accepted finding that non-nominal antecedents are only available for pronominal reference right after being uttered (Passonneau, 1991; Webber, 1991; Fraurud, 1992; Asher, 1993; Poesio and Modjeska, 2002; Gundel et al., 2003, and others). The richer semantics of SNs when compared to pronouns could in part be responsible for this (Kolhatkar, 2015), but this leaves open the question of how possible antecedents are handled in the discourse model in the first place.

If non-nominal antecedent candidates are kept in a stack-like structure like NPs in Centering (Grosz et al., 1995; Poesio and Modjeska, 2002), we need to assume that the interlocutors permanently keep track of one or two highly accessible non-nominal antecedents that are available for pronouns, and additionally, a considerable number of less recent (and thus less salient) antecedents, which can be referred to by SNs only. However, if that is the case, it seems odd that such antecedents are generally referred to by SNs accompanied by the demonstrative determiner *this* (cf. examples 4–5) which indicates a *high* accessibility of the antecedent (Gundel et al., 2003; Poesio and Modjeska, 2002). Thus, global SNs refer to pieces of information that have not necessarily been uttered recently, yet are highly accessible, both of which are attributes of DTs.

Instead of analyzing the underlined portions of the discourses in examples (4–5) as instances of long-distance antecedents, we can view them as stretches of text that function as *topic descriptions*. Even though the subsequent SNs then refer to the very topics described by the apparent antecedents, they do so only as a result of a meta-linguistic interpretation of the discourse without reflecting an underlying, direct anaphoric process. That is, annotators arrive at these antecedents by first, understanding the SN as referring to a DT; second, identifying the intended topic; and third, scanning the text for a suitable paraphrase of it.

The process of interpreting global SNs thus gives the impression of being a combination of semantic-pragmatic tasks like word sense disambiguation, topic detection, question answering, key phrase extraction, etc. Bearing little resemblance to the much more compositional interpretation of local SNs, it appears that entirely different criteria need to be applied to global SNs, as factors previously assumed to play a role for the resolution of SNs as a whole, like the syntactic shape of the antecedent, distance to the SN, and local discourse structure, seem to be primarily relevant for the subset of local SNs.

3.2 Evidence for Global Shell Nouns

So far, global and local SNs have not been investigated separately, which makes it difficult to assess the impact of global SNs on the resolution of the class as a whole. However, by analyzing the previously published data by Kolhatkar et al. (2013) and Kolhatkar (2015) it can be shown that global SNs tend to be harder to annotate and to resolve, which is in line with the idea that the processes underlying global SN resolution are more complex than for local SNs.

As SNs need to be semantically compatible with their antecedents (Schmid, 2000), we can expect nouns with a “topic-like” semantics to be more likely to refer to DTs than others. Furthermore, given that newspaper articles tend to be about everyday and social *issues* and *questions* as well as political and court *decisions*, it is possible to divide the selection of SNs in Kolhatkar et al. (2013) into a local (*fact*,

	<i>Fact</i>	<i>Reason</i>	<i>Issue</i>	<i>Decision</i>	<i>Question</i>	<i>Possibility</i>	<i>All</i>
$c < .5$	8%	8%	36%	21%	13%	7%	16%
$.5 \leq c < .6$	6%	6%	13%	8%	7%	5%	8%
$.6 \leq c < .8$	24%	25%	31%	31%	22%	27%	27%
$.8 \leq c < 1.$	22%	23%	11%	14%	19%	25%	18%
$c = 1.$	40%	38%	9%	26%	39%	36%	31%
Average c	.83	.82	.61	.72	.80	.83	.76

Table 1: Results of the crowd-sourcing annotation by (Kolhatkar et al., 2013, p. 116)

reason, *possibility*) and a global subset (*issue*, *question*, *decision*), with the latter being much more likely to refer to topics.

Among the annotation experiments conducted by Kolhatkar et al. (2013), the most important for our purposes is the task where crowd-workers were asked to select the host sentence of the antecedent without specifying an exact constituent. If topics are represented by spans of text that can occur anywhere in the text, are more difficult to pinpoint, more likely to be discontinuous, and potentially left implicit, the source sentences of topic descriptions can be expected to be more difficult to locate than the antecedents of local SNs, i.e., a better performance can be expected for the nouns *fact*, *reason*, and *possibility*, than for *issue*, *question*, and *decision*.

Kolhatkar et al. (2013) report confidence levels as returned by the CrowdFlower³ platform they used for the task (see Table 1). A low confidence means less agreement, so it is obvious that annotators agreed the least for the three nouns with a topic-like meaning. In light of these results, Kolhatkar et al. (2013, p. 116) recognize a special status of *issue* and *decision*, noting they “had a large number of low-confidence ($c < 0.5$) instances, bringing in the question of reliability of antecedent annotation of these nouns”. Further explanation is not provided, and the data is not publicly available, but it seems likely that the lower annotation confidence for topic-like SNs has to do with global references.

Additional evidence for a difference in behavior between global and local SNs can be drawn from the distance to their respective antecedents, which can be expected to be longer for the former. Varada Kolhatkar kindly provided a data set from her study (Kolhatkar, 2015) consisting of a collection of antecedent sentences and SN strings alongside a link to their respective source articles in the NYT corpus. By means of these data, average distances from the SNs to their antecedents (ignoring cataphoric instances) can be calculated as 1.3 for *issue* ($n=265$); 1.2 for both *decision* ($n=343$) and *question* ($n=376$); 1 for *possibility* ($n=268$); and 0.8 for both *fact* ($n=436$) and *reason* ($n=412$). As the data did not include the exact positions of the SNs and the antecedents in the source articles, their locations had to be determined heuristically, hence the results may not be entirely accurate, but there is a clear tendency that global nouns exhibit higher average distances.

This is further supported by the baseline algorithm presented by Kolhatkar (2015, p. 121) that selects as antecedent the sentence preceding the host sentence of the SN which gives us an idea of the distribution of antecedent distances. Table 2 shows the ratio of antecedents with a distance of 1 according to this data. As expected, the three global nouns have the lowest percentage of adjacent antecedents. While this could also indicate a higher number of sentence-internal antecedents, it is likely that long-distance antecedents are responsible for the lower number of nearby antecedents.

Kolhatkar (2015, p. 119) notes that the nouns *issue* and *decision* are also idiosyncratic with respect to the distribution of the syntactic types of their antecedents, exhibiting less sentences and clauses, and more NPs and VPs. According to her, this is a result of the nouns being more flexible than others regarding the types of abstract objects they can refer to, but it could also be an indicator of the meta-linguistic nature of global SN interpretation outlined above, following the intuition that annotators who look for topic descriptions based on semantic considerations – as opposed to the more syntactically driven task of finding proper antecedents – will tend to select a higher number of syntactically atypical antecedents.

³<http://crowdfLOWER.com/>

	<i>reason</i>	<i>fact</i>	<i>possibility</i>	<i>question</i>	<i>decision</i>	<i>issue</i>
Antecedent in PS	44%	40%	34%	25%	21%	19%

Table 2: Percentage of antecedents found in the sentence preceding the SN (Kolhatkar, 2015, p. 121)

Finally, the results of the resolution systems reported by Kolhatkar et al. (2013) and Kolhatkar (2015) consistently show a lower performance for the nouns *issue* and *decision*, which is hardly surprising as difficulties in the annotation can be expected to propagate to the final resolution.

3.3 Annotation Experiment

In order to gain a more direct (if preliminary) assessment of how frequent global SNs are, a small annotation experiment has been conducted. To this end, four German SNs have been selected: *Entscheidung* (‘decision’) and *Angelegenheit* (‘issue’) as examples of topic-like SNs, and *Tatsache* (‘fact’) and *Problem* (‘problem’) as local SNs. For each noun, 20 articles have been collected from the German online newspaper *Die Zeit*, each containing one instance of the noun accompanied by the demonstrative determiner *diese(r)* (‘this’).

After receiving an introduction about SNs and the idea of global reference, two German native speakers majoring in linguistics were instructed to carefully read the articles and decide whether they think the SNs in question were GLOBAL or LOCAL. In addition, they were allowed to mark the nouns as COREFERENTIAL if they thought it was part of an anaphoric chain and did not directly refer to a non-nominal antecedent, or UNCLEAR if they could not decide for sure. One of the articles (containing an instance of the German pendant of *fact*) had to be dismissed, resulting in a total number of 79 items.

The two analysts agreed in 52 cases (65.8%), 43 (54.4%) of which they evaluated as LOCAL, and 7 (8.9%) as GLOBAL. The global items consisted of one instance of *decision*, two instances of *problem*, and 4 instances of *issue*, i.e., the latter has been agreed by the analysts to be global 20% of the time.

While only looking at a small sample, this study suggests that references to DTs are a considerable factor at least for some SNs, particularly the noun *issue*. The experiment also revealed that a simple global/local distinction is too coarse as SNs can be used to refer to local DTs, i.e., DTs of discourse sections, which caused uncertainty among the annotators who reported after the study that they were sometimes unsure whether references to topics of the host paragraph of the SN had to be annotated as GLOBAL or LOCAL. Furthermore, the annotators often found the nouns *fact*, *decision*, and *problem* easy to identify as local instances with syntactically prominent antecedents in their close vicinity, while *issue* was generally hard to annotate due to its vague semantics.

4 Discussion

The fact that the global/local distinction has not been acknowledged in previous work appears to be a major cause of misunderstanding of SNs which led to flaws in annotation and resolution approaches. The majority of the criteria previously employed to resolve SNs are only applicable to local SNs, whereas the resolution of global SNs is a novel, separate task that needs additional steps to be introduced into the resolution pipeline. This is particularly apparent in the factor of distance between the SN and its antecedent.

Although distance is very important for evaluating the accessibility of local antecedents, it is in principle irrelevant for the resolution of global SNs, as spans of text describing the DT can occur virtually anywhere in the text or even be left implicit, which puts both resolution systems and human annotators in an awkward position: they are expected to look for an antecedent where there is none, resulting in a high uncertainty and variance across annotations. Furthermore, it appears that the anaphoric range of SNs has been overestimated as a result of topic references being analyzed as long-distance antecedents. The data from previous studies suggest that local SNs will only rarely refer to an antecedent further away than one sentence, and any antecedent that occurs further away than that is either still accessible because of properties of the local discourse structure (Webber, 1991; Asher, 2008) or it is not an antecedent in the sense of a reflection of an underlying anaphoric process.

Regarding annotation, it needs to be asked whether a classic annotation setting is appropriate for global SNs. Annotators will be biased to annotate *something* in the vicinity of the target noun even if the available context does not provide a stretch of text that perfectly matches their understanding of the noun. In order to avoid such bias, it might be preferable to approach SNs by means of a more open question answering task, asking annotators to paraphrase what they think the target noun means in context, and subsequently comparing their answers with each other and with the source text. Ideally, such a strategy would provide a more accurate picture of how SNs are understood. The evaluation of free text answers is very difficult and not well researched, though, but it bears resemblance, e.g., to the evaluation of answers to reading comprehension tasks (Hahn and Meurers, 2012), which could be adapted to SN research.

Due to their semantics, some SNs lend themselves more readily to a global reading than others. Such tendencies are both domain- and text-specific, but it is fair to assume that the nouns *issue*, and, to a lesser extent, *decision*, are likely to refer to the overall topic in the domain of newspaper articles. The focus of previous work on instances of such nouns seems to have caused an over-representation of topic references in the data. This is further amplified by limiting the considerations to instances of nouns with the demonstrative determiner *this*, which is associated with a high accessibility of the antecedent (Kolhatkar, 2015). As shown by the data by Simonjetz and Roussel (2016), SNs with an accompanying *that* tend to refer to closer antecedents (average distance=0.84 sentences; n=18) than SNs with *this* (d=1.13; n=23), indicating that the former might be correlated with local, and the latter with global SNs. As nouns with a less topic-like meaning like *fact* proved to be much easier to annotate and resolve, SN resolution might be better off focusing on such nouns first, expanding the data to determiners other than *this* to prevent a bias towards topic references.

Assuming that topic referents compete with locally available, explicit antecedents, the factors that are helpful for the resolution of local SNs are probably also helpful for identifying global SNs, following the intuition that if there is no suitable antecedent nearby, the SN derives its information elsewhere. Thus, what both human annotators and resolution systems are missing is the option to disregard the available local antecedents in favor of a discourse topic. As a tentative strategy to account for such an option, the set of antecedent candidates could be extended by one or more pseudo-antecedents representing the discourse topic(s). The candidate set would then consist of antecedents extracted from the close proximity of the SN and a number of topic descriptions, making for a considerably reduced search space and leaving human annotators and automatic resolvers free to choose among explicit antecedents and topics as equally viable options. Future work will need to explore the details of such an approach, addressing questions such as how to extract topic information, how to generate antecedent candidates from topic(s), how to treat features such as distance and phrase type for such antecedents, etc.

5 Conclusion

Quantitative approaches to SNs have proved difficult both in terms of annotation and automatic resolution, and part of this difficulty seems to be due to misconceptions about the interplay between SNs and their context. The prevalent assumption that the occurrence of a SN “*typically* [emphasis added] involves a full-fledged clausal antecedent” (Marasovic et al., 2017, p. 222) seems to have caused more harm than good by raising wrong expectations about SN antecedents, which apparently caused references to DTs to be mistaken for long-distance antecedents. The evidence from theoretical work on anaphoric reference to non-nominal antecedents, practical approaches to SNs, and the annotation experiment outlined above, suggest that some SNs, notably *issue* and *decision* – which much of previous work on SN resolution is based on – are frequently used to refer to DTs that may or may not be explicitly present in the discourse. Crucially, the detection of spans of text that match a topic description is an entirely different task than the resolution of local SNs. Future research will need to find ways to deal with global SNs, topic antecedents, and other implicit sources of abstract referents, in order to make manual annotation and automatic resolution systems more reliable.

References

- Mira Ariel. 1988. Referring and Accessibility. *Journal of linguistics*, 24(1):65–87.
- Nicholas Asher and Alex Lascarides. 2003. *Logics of conversation*. Cambridge University Press, Cambridge.
- Nicholas Asher. 1993. *Reference to Abstract Objects in Discourse*. Kluwer, Dordrecht.
- Nicholas Asher. 2000. Events, facts, propositions, and evolutive anaphora. In James Higginbotham, Fabio Pianesi, and Achille C. Varzi, editors, *Speaking of events*, pages 123–150. Oxford University Press, New York.
- Nicholas Asher. 2004. Discourse topic. *Theoretical Linguistics*, 30(2–3):163–201.
- Nicholas Asher. 2008. Troubles on the right frontier. In Anton Benz and Peter Kühnlein, editors, *Constraints in Discourse*, pages 29–52. John Benjamins Publishing Company, Amsterdam.
- Stefanie Dipper and Heike Zinsmeister. 2010. Towards a standard for annotating abstract anaphora. *Proceedings of the LREC 2010 workshop on Language Resources and Language Technology Standards*, pages 54–59.
- Stefanie Dipper and Heike Zinsmeister. 2012. Annotating abstract anaphora. *Language Resources and Evaluation*, 46(1):37–52.
- John Flowerdew and Richard W. Forest. 2015. *Signalling nouns in English*. Cambridge University Press, Cambridge.
- John Flowerdew. 2003. Signalling nouns in discourse. *English for Specific Purposes*, 22(4):329–346.
- Gill Francis. 1994. Labelling discourse: An aspect of nominal-group lexical cohesion. In M. Coulthart, editor, *Advances in written text analysis*, pages 83–101, London. Routledge.
- Kari Fraurud. 1992. Situation reference: What does “it” refer to? In *Processing noun phrases in natural discourse*, Department of Linguistics, Stockholm University. PhD thesis.
- Barbara J. Grosz, Scott Weinstein, and Aravind K. Joshi. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.
- Jeanette K Gundel, Michael Hegarty, and Kaja Borthen. 2003. Cognitive status, information structure, and pronominal reference to clausally introduced entities. *Journal of Logic, Language and Information*, 12(3):281–299.
- Michael Hahn and Detmar Meurers. 2012. Evaluating the meaning of answers to reading comprehension questions: A semantics-based approach. In *Proceedings of the Seventh Workshop on Building Educational Applications Using NLP*, pages 326–336.
- Michael A. K. Halliday and Ruqaiya Hasan. 1976. *Cohesion in english*. Longman, London.
- Roz Ivanič. 1991. Nouns in search of a context: A study of nouns with both open- and closed-system characteristics. *IRAL – International Review of Applied Linguistics in Language Teaching*, 29(2):93–114.
- Varada Kolhatkar and Graeme Hirst. 2012. Resolving “this-issue” anaphora. In *Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing*, pages 1255–1265, Jeju Island, Korea.
- Varada Kolhatkar and Graeme Hirst. 2014. Resolving shell nouns. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 499–510, Doha, Qatar.
- Varada Kolhatkar, Heike Zinsmeister, and Graeme Hirst. 2013. Annotating anaphoric shell nouns with their antecedents. In *Proceedings of the 7th Linguistic Annotation Workshop & Interoperability with Discourse*, pages 112–121, Sofia, Bulgaria.
- Varada Kolhatkar, Adam Roussel, Stefanie Dipper, and Heike Zinsmeister. 2018. Anaphora with non-nominal antecedents in computational linguistics: A survey. *Computational Linguistics*, 44(3):547–612.
- Varada Kolhatkar. 2015. *Resolving Shell Nouns*. Ph.D. thesis, University of Toronto.
- Klaus Krippendorff. 2013. *Content analysis: An introduction to its methodology*. Sage, Thousand Oaks, CA, USA, 3rd edition.

- Ana Marasovic, Leo Born, Juri Opitz, and Anette Frank. 2017. A mention-ranking model for abstract anaphora resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 221–232.
- Rebecca J. Passonneau. 1991. Some facts about centers, indexicals, and demonstratives. In *29th Annual Meeting of the Association for Computational Linguistics*, pages 63–70, Berkeley, California, USA. Association for Computational Linguistics.
- Massimo Poesio and Ron Artstein. 2008. Anaphoric annotation in the ARRAU corpus. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*.
- Massimo Poesio and Natalia N. Modjeska. 2002. The THIS-NPs hypothesis: A corpus-based investigation. *Proceedings of the 4th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC 2002)*, pages 157–162.
- Massimo Poesio, Roland Stuckardt, and Yannick Versley. 2016. *Anaphora resolution: Algorithms, resources, and applications*. Springer, Heidelberg, Berlin.
- Craige Roberts. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5.
- Adam Roussel, Stefanie Dipper, Sarah Jablotschkin, and Heike Zinsmeister. 2018. Towards the automatic resolution of anaphora with non-nominal antecedents: Insights from annotation. In *Proceedings of the 14th Conference on Natural Language Processing (KONVENS 2018)*, pages 178–191, Vienna, Austria.
- Adam Roussel. 2018. Detecting and resolving shell nouns in German. In *Proceedings of the Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 61–67, New Orleans, Louisiana.
- Hans-Jörg Schmid. 1997. Constant and ephemeral hypostatization. In Bernard Caron, editor, *Proceedings of the 16th International Congress of Linguists*, Paris. Elsevier.
- Hans Jörg Schmid. 2000. *English abstract nouns as conceptual shells: From corpus to cognition*. De Gruyter, Berlin, New York.
- Hans-Jörg Schmid. 2018. Shell nouns in English: A personal roundup. *Caplletra*, 64:109–128.
- Fabian Simonjetz and Adam Roussel. 2016. Crosslinguistic annotation of German and English Shell Noun complexes. In *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*, pages 265–278.
- Zeno Vendler. 1967. *Linguistics in philosophy*. Cornell University Press, Ithaca, NY, USA.
- Richard Watson Todd. 2016. *Discourse topics*, volume 269 of *Pragmatics & Beyond New Series*. John Benjamins Publishing Company, Amsterdam.
- Bonnie Lynn Webber. 1988. Discourse deixis: Reference to discourse segments. In *Proceedings of the 26th Annual Meeting on Association for Computational Linguistics*, ACL '88, pages 113–122, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Bonnie Lynn Webber. 1991. Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes*, 6(2):107–135.