Natural Language Engineering **4** (4): 345–361. Printed in the United Kingdom

© 1998 Cambridge University Press

Understanding pronominal references to quantified expressions†

ELISE H. TURNER
Computer Science Department, University of Maine, Orono, ME 04469, USA
e-mail: eht@umcs.maine.edu

COLLETTE M. MATTHIAS
Sanders, A Lockheed Martin Co., PO Box 868, MER15-2403, Nashua, NH 03061, USA
e-mail: collette.m.matthias@lmco.com

(*Received 9 October 1997; Revised 25 August 1998*)

Abstract

Quantified expressions present an interesting case for understanding pronominal reference because the quantifier which appears in the expression affects how the pronoun will be understood. The quantifier determines which sets will be placed in focus and predicts the relation between the sentence containing the quantifier and the following sentence (Moxey and Sanford 1993a). This information is essential to understanding the pronominal reference. In this paper, we discuss how a natural language processing system can take advantage of this information to understand pronominal references to quantified expressions.

1 Introduction

Understanding pronominal references to quantified expressions is an unusual case for natural language systems because the quantifier that is used in the initial referring expression impacts how the pronoun will be understood. Consider the following sentence pairs, based on examples from Moxey and Sanford (1993a):

A *Few* of the children saw the gorilla.
B They saw the chimpanzee beside her.

C *Many* of the children saw the gorilla.
D They saw the chimpanzee beside her.

In Sentence 1, 'they' is understood to refer to the children who *did not* see the gorilla, and the sentence gives an explanation for why these children did not see the gorilla – they were looking at the chimpanzee, instead. Most natural language

† This material is based upon work supported by the National Science Foundation under Grant No. IRI-9696055. The authors wish to thank Rich Cullingford, Roy Turner, Roger Kibble, James Lawton, Prabha Ramakrishnan, John Phelps and Tom Roper. The authors also wish to thank the anonymous reviewers for their helpful comments.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

systems would not find this interpretation. Instead, 'they' would be understood as referring to the children who *did* see the gorilla, the set that is explicitly referenced in the sentence. With this interpretation of 'they', Sentence 1 would be understood as describing what the children did after they saw the chimpanzee. Although this interpretation would not be preferred for the first pair of sentences, it would be the preferred interpretation for the second pair which differs only in the quantifier used. However, most natural language systems would not make this distinction because most do not take into account important information associated with the quantifiers.

Moxey and Sanford (1993a) identify two types of information which are associated with quantifiers and which must be taken into account to properly process sentences like those above:

The sets which are placed in focus by the quantifier. Moxey and Sanford, like many computational linguists (e.g. Grosz 1977; Sidner 1983; Reichman 1985) use 'focus' to mean *attentional focus*. Entities are placed in focus because they dominate the processing of an utterance or are highlighted in the discourse in some way. The notion of focus is important for understanding pronominal reference because it distinguishes the entities that can be referred to by a pronoun.

Most natural language processing systems require, at a minimum, that the entity be explicitly referenced in the preceding discourse to be placed in focus. However, certain quantifiers have the ability to bring other sets into focus. For example, 'few' in Sentence 1 places the children who did not see the gorilla in focus, allowing them to be the referent of 'they' in Sentence 1.

Predictions for the relation between the sentence containing the quantifier and the following sentence. The ability to find a coherence relation is an important test of a potential referent of a referring expression (Hobbs 1979). In natural language systems, these coherence relations are often suggested by knowledge of discourse structure (e.g. Mann, Moore and Levin 1977; McKeown, 1985; Grosz and Sidner 1986). However, predicted relationships associated with quantifiers also provide an important source of information about how sentences are related. In addition, the predicted relations can direct the reasoning required to select between the potential referents. For example, 'few' predicts that the following sentence will provide the reason why so few children were specified in the first sentence. We find that relation by understanding the sentence pair to mean "Few children saw the gorilla because they were looking at the chimpanzee beside her, instead". Once we verify that the predicted relation can hold, it is easy to see that 'they' must refer to the children who did not see the gorilla.

The paper begins with an overview of the work of Moxey and Sanford in section 2. Their studies suggest that the information described above is part of the semantics of quantifiers. They also identify the focus sets and predicted relations for several quantifiers. In section 3, we discuss how these findings can be integrated into a natural language processing system. We conclude in section 4 with a discussion of the current state of our project and future work.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

2 Knowledge associated with quantifiers

Moxey and Sanford (1993a) argue that instead of denoting a specific number or proportion, quantifiers control the inferences that will be made by the reader or listener. Some of these inferences pertain to the size of the set being referenced. They allow the reader/listener to understand what is important about the size of the set without actually computing its size.

Different quantifiers also emphasize different aspects of the size of the set. This emphasis is captured by associated inferences which place sets in focus and predict the relation between the sentence containing the quantifier and the following sentence. Quantifiers can be divided into two classes. *Positive quantifiers* are monotone increasing and implicate 'not all'. 'Some', 'many' and 'a few' are all examples of positive quantifiers. *Negative quantifiers* are monotone decreasing and implicate 'not none'. 'Few', 'hardly any' and 'not quite all' are examples of negative quantifiers.[1] Negative quantifiers make assertions about the smallness of a set relative to prior expectations (Moxey and Sanford 1993b) and, thus, bring into focus the set of entities for which the predicate is not true.

In this paper, we are concerned only with the inferences which place sets in focus. These can be incorporated in a natural language system to find the correct referents for pronominal references to quantified expressions, regardless of the how the referent of the quantified expression is represented. Consequently, we will not discuss the rich literature on reference and quantification in linguistics, philosophy and psychology that is related to Moxey and Sanford's general approach. We will instead mention two methods for representing quantifiers that are often used in computer systems and describe in section 2.1 how they can be used to represent the sets which are placed in focus.

Moxey and Sanford's interpretation of quantified expressions differs significantly from much of the work in which quantified expressions are represented by formal semantics. For example, Kamp (1981) and Heim (1982) represent quantified expressions in terms of restricted sets. However, these treatments do not appear to allow pronominal references to the set of entities for which the predicate is not true. The notion of restricted sets can be used in computer systems (Allen 1995), even when those systems do not follow Kamp and Heim when finding referents to anaphoric expressions. In fact, as discussed below, some mechanism for identifying the objects which will or will not be in the set will be necessary to exploit Moxey and Sanford's insights about understanding pronominal references to quantified expressions. Moxey and Sanford's approach is also a departure from previous work in psycholinguistics and psychometrics (e.g. Holyoak and Glass 1978; Hartley, Trueman and Rogers 1984; Wright, Gaskell and O'Muircheartaigh 1994) which suggests that quantifiers can be translated into magnitudes, scales, or ranges of numbers. This has been adopted in computer systems by representing vague quantifiers as fuzzy sets (Zadeh 1965), or, more naively, as magnitudes or percentages. Although Moxey and Sanford's semantics for quantifiers rejects this approach, systems which use it can still exploit Moxey and Sanford's insights about focus.

[1] See Moxey and Sanford (1993a) for more details about this classification.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

2.1 Identifying sets in focus

Moxey and Sanford identify three sets which may be placed in focus by quantified expressions. As negative quantifiers emphasize the smallness of a set, they bring attention to the items that are not included in the set. Because the items that were not included are given this attention, they are placed in focus and are made available as potential referents of pronouns.[2] Positive quantifiers emphasize the largeness of the set, and, so, place in focus only the items included in that set. To illustrate the sets that can be placed in focus, we return to Sentence 1, "Few of the children saw the gorilla". 'Few' is a negative quantifier, so three sets are placed in focus:

1 The set of all children.
2 The set of children who saw the gorilla.
3 The set of children who did not see the gorilla.

The first set, called the *wholeset*, comes into focus due to the explicit mention of children within the quantified expression. This is not dependent on the quantifier. The second set is also explicitly referenced in the text. It is the subset of the wholeset for which the predicate is true. The quantifier causes these children to be distinguished from other children in the wholeset. This set is called the *refset*. The refset is placed in focus by both positive and negative quantifiers. The third set is not explicitly referenced. This set is called the *compset* because it is the complement of the refset. It is the subset of the wholeset for which the predicate is false. The compset is placed in focus only by negative quantifiers.

Not all researchers agree that the compset is being referenced (e.g. Kamp and Reyle 1993; Corblin 1996). Instead they argue that what appears to be the compset is actually a generalization or a set of typical members. (A more detailed discussion of these arguments can be found in Kibble (1997).) However, Moxey and Sanford have experimental evidence that supports reference to the compset. When subjects were asked to identify the referent of 'they' in sentences which they wrote, they selected the compset instead of the wholeset (e.g. *all of the children*) or a generalized set (e.g. *children, in general*) (Moxey and Sanford 1993a; Sanford, Moxey and Paterson 1996). Also, subjects selected the compset as the referent of 'they' consistently across a number of quantifiers, independent of the proportions that they denote. In cases of quantifiers which denote a high proportion, the general set would correspond to the refset. For example, in "Not quite all children saw the gorilla", most children, and so the typical child, would have seen the gorilla. Since the subjects selected the compset even when quantifiers denoted high proportions, this experimental evidence suggests the existence of the compset (Sanford *et al*. 1996).

The representation of the compset and refset in any particular system will necessarily depend on the knowledge representation used in that system. However, the representation of the quantified expression cannot simply give the approximate

[2] Different systems use different methods for determining which sets are in focus and how they will be made available as referents of pronouns. Here, we simply summarize Moxey and Sanford's argument for placing the implicitly referenced set in focus. In section 3, we discuss how this information might be used in a natural language processing system.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

number of items of a particular type, as may be the case in representations based on numeric scales. For instance, in our example, "Few children" cannot simply be represented as 'less than 20 children'. Instead, it is critical that the set's membership be identified using information in the predicate of the sentence. In our example, we might represent that the number of members in the set is less than twenty, but we must also represent that these are children *who saw the gorilla*. The representation must also provide a mechanism for forming the complement of this identifying information so the compset can be represented. For example, this information might be captured in a slot describing typical members in a system using slot-filler representation (Cullingford 1986) or as restricted quantification in a system using predicate calculus (Allen 1995). Recent work by Kibble (1997) suggests definitions for monotonically decreasing quantifiers which support reference to the compset and which could be adopted by systems using predicate calculus.

2.2 Predictions for relations

Because different quantifiers are associated with different patterns of inferences, they set the stage differently for the *continuation sentences* that follow them. For example, when 'few' brings the compset into focus, it leads the reader/listener to expect to be told why many entities were not included in the refset. The *continuation type* describes the relation between the sentences in terms of the information given about one of the sets in focus. Moxey and Sanford identify patterns of continuation associated with several quantifiers. Although other continuation types may be allowed, these are the predicted types. Their work also suggests that there is a dominant, and sometimes a secondary, continuation type for each quantifier.

Moxey and Sanford identify four continuation types which obtain between quantifiers and continuation sentences:

WHN: The continuation sentence describes *what happened next* to the refset. This is the dominant continuation type for 'many' and 'a few'. It is also the secondary continuation type for 'few'.

RW: The continuation sentence describes the *reason why* the predicate in the first sentence is true of the refset. This is the secondary continuation type for 'a few' and 'many'.

RWN: The continuation sentence describes the *reason why* the predicate is *not* true of the compset. This is the dominant continuation type of all negative quantifiers studied, including 'few".

CONS: The continuation sentence gives a *consequence* of the proportion denoted by the quantifier. Although seen as a continuation type, it is neither the dominant nor the secondary continuation type of any of the quantifiers studied.

In the studies which suggested the continuation patterns of quantifiers, subjects were asked to complete continuation sentences beginning with 'they'. This forced subjects to write continuation sentences that were relevant to the quantified expression (Moxey and Sanford 1993a). Since we are interested in understanding pronominal

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

references to quantified expressions, predictions based on these continuation patterns are appropriate for our work.

The patterns of continuation guide reasoning in two ways that are important for natural language systems. First, they provide predictions about the relation that will hold between a sentence and its continuation, allowing the system to test the relations in order of dominance. Secondly, they predispose the understander to find the predicted relation. In cases where several relations might hold, the understander can make inferences that are appropriate for supporting the dominant relation. For example, natural language systems which ignore information associated with the quantifier would interpret Sentence B, "They saw the chimpanzee beside her", as describing what happened after the children saw the gorilla. However, because RWN is the dominant continuation type for 'few', our system would consider that relation first, making the inferences needed to understand how seeing the chimpanzee could interfere with seeing the gorilla.

3 Extending a natural language system to exploit information associated with quantifiers

In this section, we will describe how an existing natural language system can be extended to understand references to quantified expressions. We have implemented these extensions in the NLPTool (Cullingford 1986) as part of our interface to Autonomous Underwater Vehicles (AUVs). Details of the implementation of our extensions to the NLPTool can be found in Matthias (1997).

AUVs are unmanned, untethered vehicles that are being developed for underwater scientific exploration. Our ultimate goal is to allow scientists in their laboratories to interact with AUVs as they perform their mission. Initially, we expect scientists to use quantifiers when giving instructions about samples. As the AUVs' reasoning mechanisms and interface become more sophisticated, the scientists and AUVs will work together to plan missions. In these cases, we expect to see an increase in negative quantifiers as well as RW continuation sentences for positive quantifiers as the interlocutors reason about and seek to explain their domain. Our current natural language interface functions mainly to give commands to simulated AUVs (Turner, Blidberg, Fox and Turner 1991). The example sentences in this paper are about AUV missions so that the continuation types can be verified using the AUVs' mission planners.

3.1 Identifying potential referents

In this section, we discuss several methods for identifying potential referents that are used in natural language systems and describe how they might be extended to take Moxey and Sanford's results into account. To illustrate this discussion, we return to our first example sentence pair:

A *Few* of the children saw the gorilla.
B They saw the chimpanzee beside her.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

To handle a pronominal reference to a quantified expression, the sets which the quantifier places in focus, and only those sets, must be recognized as potential referents of pronouns. Some natural language systems use a model of focus to identify potential referents of pronouns. Other systems do not have a model of focus but, instead, rely on some other criteria to suggest referents for any anaphoric reference. In both types of systems, usually only those entities which have been explicitly referenced in the discourse are available as referents of pronouns. Consequently, each must be extended so the implicitly referenced compset is available to serve as the referent. However, it is important that the model not commit to either the compset or the refset when both are available because either may be the referent of the pronoun.

Systems which model focus can be made to handle quantified expressions simply by extending the rules which place entities in focus. Because focus is meant to reflect the amount of attention an entity receives, and, consequently, the ease with which the understander can make inferences about that entity (Grosz, Joshi and Weinstein 1995), many systems classify (e.g. Reichman 1985) or rank (e.g. Grosz, Joshi and Weinstein 1995) discourse entities by level of focus. Only entities with a high level of focus can be referred to with a pronoun. In systems which classify the discourse entities, the compset and the refset must be placed at the highest level of focus when the referring expression contains a negative quantifier.

For example, consider the widely-used centering model of Grosz, Joshi and Weinstein (1995). In this model, entities are ranked according to how they are realized in an utterance. These are called the *forward-looking centers* of the utterance. Grosz, Joshi and Weinstein (1995) suggest that the grammatical role and surface position are two important factors in this ranking. The *backward-looking center* of an utterance connects that utterance to the previous utterance. The ranking of the forward-looking centers of one sentence is a major determinant of the backward-looking center of the next. The higher the rank of some entity in the forward-looking centers, the more likely it is to be the backward-looking center of the next utterance. And, the backward-looking center of an utterance will be the highest-ranked entity appearing in the sentence. To find the referent of the pronoun which realizes the backward-looking center, a system can order the potential referents according to the rules of centering and then rule out potential referents by reasoning about the sentence (Brennan, Friedman and Pollard 1987). So, when the backward-looking center of an utterance is pronominalized, the referent of the pronoun will be the highest-ranked forward-looking center which could serve as the referent of the pronoun (based on agreement and reasoning about the possible referents).

Grosz, Joshi and Weinstein (1995) specify grammatical role as the major determinant of ranking of the forward-looking centers. Consequently, systems using this ranking method, without taking Moxey and Sanford's results into account, would give the following ranking of the forward-looking centers of Sentence A (from highest to lowest): *the children who saw the gorilla* (the refset), *the gorilla*, *all children* (the wholeset). 'They' is the backward-looking center of Sentence B. The refset would be selected as the referent of 'they' because it is the highest-ranked forward-looking center which could serve as the referent of the pronoun. This would not agree with our interpretation of 'they' as the compset.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

However, Grosz, Joshi and Weinstein (1995) do not insist that forward-looking centers be ranked in this way, and, in fact, include lexical semantics as an additional factor which may affect the ranking. This allows us to easily incorporate Moxey and Sanford's results. Moxey and Sanford argue that the compset is brought into focus as part of the semantics of the quantifier. So, when a negative quantifier is seen, both the compset and refset can be placed in focus. Since Moxey and Sanford suggest that negative quantifiers place more attention on the compset, it should be ranked higher than the refset. Other referring expressions which evoke entities that are not explicitly referenced may not have these entities so highly ranked, unless warranted by their semantics. After including Moxey and Sanford's results, the forward-looking centers of Sentence A would be: the compset, the refset, *the gorilla*, and the wholeset. The system will prefer the compset. So, in our example the compset will be selected as the referent because it is the highest ranked, it agrees with the pronoun, and this interpretation is not eliminated by reasoning about the world. If this interpretation were eliminated, the refset would be tried next. For example, if the continuation sentence were "They laughed when she ate a banana", the compset would not be seen as a reasonable interpretation and the refset would be found as the referent of 'they'. The reasoning to test potential referents can be guided by continuation types, as described below.

Systems which do not rely on focus must also make both the compset and refset accessible for negative quantifiers. These systems include the NLPTool as well as those using RAP (Resolution of Anaphora Procedure) (Lappin and Leass 1994), and history lists (Allen 1995). All maintain a collection of entities that have been referenced in the discourse and find referents for pronouns by searching the collection to find a referent which agrees with the pronoun in number and gender. Different systems use different criteria for selecting the referent: RAP uses a combination of salience factors, history lists rely on recency, and the NLPTool returns all entities which agree with the reference so that the actual referent can be selected by some other mechanism. If a history list were used to process our example, after Sentence A the history list would include (in order of recency): *the gorilla*, *the children who saw the gorilla* (the refset), and *all the children* (the wholeset). The refset would be selected as the referent for 'they', since it is the first entity in the list to agree with the pronoun.

The example list above includes only explicitly referenced entities, although there are ways to include implicitly referenced entities as well. This leaves open the possibility of including the compset. However, their selection criteria predispose these mechanisms to find entities that are explicitly referenced. For example, one way to include implicitly referenced entities in a history list is to add related entities at the time the explicitly referenced item is added. Since history lists use recency as their selection criterion, we can avoid allowing implicitly referenced objects to be selected as referents to pronouns simply by making them appear to be less recent than the explicitly referenced object. For example, if the history list is implemented with a stack, the implicitly referenced entity can be placed beneath the explicitly referenced entity. It would be easy to specify that the compset should be added whenever a quantified expression is to be understood. For Sentence A, the history

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

list would contain: *the gorilla*, the refset, the compset, and the wholeset. Here, the compset would not be found as the referent for 'they' because it is not the first item on the history list which agrees with the pronoun. Since there is also a legitimate interpretation of Sentence B using the earlier-occuring refset as the referent of 'they' (*The children who saw the gorilla saw the chimpanzee next.*), the refset could pass any pragmatic tests. Consequently, the compset, and the preferred interpretation, would not be considered. To avoid this problem, we must take advantage of the fact that the compset will be preferred for negative quantifiers and that tests for the dominant continuation type (usually RWN) will fail if the wrong referent has been selected. Any mechanism that selects a single referent should cause the compset to be selected first when a pronoun co-refers with a quantified expression. For example, the compset can be given the place of the most recent item in a history list. The system then must perform tests to ensure that the selected referent makes sense on pragmatic grounds. The continuation types can be used to help guide these tests.

We were fortunate that the NLPTool does not select a single referent. Instead both the compset and the refset are returned from the *nlpcontext*, the NLPTool's list of discourse entities, when a referent is sought for 'they'. Previously, only items that are explicitly referenced are ordinarily placed in the *nlpcontext*, and no implicitly referenced items are generated when a referent of a pronoun is needed. The wholeset is also added to the *nlpcontext* when it is recognized as the quantified expression is being processed. For positive and negative quantifiers, when the refset has been recognized, it is added to the *nlpcontext*. So, after Sentence A, the *nlpcontext* would contain *the gorilla*, the refset, and the wholeset. We have extended the definitions of negative quantifiers in the NLPTool so that a representation of the compset is constructed from information about the wholeset and the refset, and it is added to the *nlpcontext*. This way, the compset is only available as the referent of a pronoun when co-referring with a negative quantifier and, so, is effectively placed in focus. The sets are marked as 'compset' or 'refset' because this information will be needed in further processing. With this extension, after Sentence A, the *nlpcontext* includes the compset in addition to *the gorilla*, the refset, and the wholeset. The refset, the compset, and the wholeset are all returned as possible referents for 'they' in Sentence B, and a single referent is selected as described below.

3.2 Selecting the proper referent

In this section, we describe how the continuation type is found in our system and how this leads to finding the proper referent. The tests for continuation types must depend on the inference mechanisms associated with the natural language system. In our work, the AUV's mission planner provides the reasoning capabilities and world knowledge required for the tests. The current mission planner, *Orca* (Turner 1994), is a schema-based reasoner which relies on generalized episodes, called *procedural schemas*, to provide the building blocks for its plans.

We have currently implemented simple tests for three continuation types: WHN, RWN, and RW. Test sentences are pairs of actions for WHN and RWN. We chose to use pairs of actions for RWN to demonstrate how a change in the quantifier

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

could change the interpretation of the sentence pairs. Because RW is a secondary continuation type for the quantifiers implemented, continuation sentences that are actions are interpreted as WHN whenever the action described in the continuation sentence could follow the action described in the first sentence. So, RW test sentences are state/goal and action/goal pairs.

3.2.1 Finding the continuation type

Our approach to finding continuation types is predisposed to find the dominant type in two ways:

Continuation types are tested in order of dominance. Continuation types, in order of dominance, are associated with each quantifier. After the continuation sentence is understood, these continuation types are tested to find the relation between the sentences.

All reasonable inferences are made to support the dominant type. Default information is assumed or replaced in order to verify the continuation type currently being tested. Information that is inferred in this way is added to the representation so that it cannot be distinguished from information explicitly stated in the text (in accordance with Fincher-Kiefer (1996)).

We see identifying the 'reasonable inferences' that should be made as an important area of on-going research. We believe that inferences should be based on shared knowledge, including stereotypes, conventions and common expertise. There should also be a limit on the amount of processing allowed to verify the continuation type. As currently implemented, our system makes inferences only about attributes of members of the set that is being considered as the referent of 'they'. Inferences can only be made when the value of the attribute is unknown or a default value. Values can only be changed if some well-known example exists for which the attribute has that value.

Tests for continuation types should rely on shared knowledge and well-known methods of reasoning about that knowledge. For example, the procedural schemas known by our AUVs represent standard scenarios for conducting a mission. Tests for continuation types check well-understood relations between schemas such as event order and preconditions.

Processing is similar for the WHN and RWN tests. A procedural schema is found that accounts for both events described in the text. The encompassing schema is found using a technique similar to the one used in PAM (Wilensky 1978). Actions described in the sentences provide indices into the low level procedural schemas which contain them. These procedures, in turn, point to the higher level procedures which include them. Each of the higher level schemas is checked until a schema which links both actions is found. We assume here that the generalized schemas of Orca will represent scenarios that will be well-known enough in the AUV community to serve as the shared knowledge needed for effective communication. After the

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

encompassing schema is found, it is checked to see if the relation being tested can be verified. WHN is verified when all preconditions in the procedural schema can be satisfied; RWN is verified when a precondition can be found which does not hold.

The RW test seeks to verify that the action or state is needed to support the goal given in the continuation sentence. [3] If the sentence containing the quantifier describes an action, the system tests the procedural schemas which contain the action. If the goal of any of these procedural schemas match the goal described in the continuation sentence, the system infers that the action was performed to achieve the goal, and RW is verified. If the first sentence containing the quantifier describes a state, the system checks if the state supports some precondition of the goal. First the system identifies procedural schemas which can achieve the goal. It then checks the preconditions of these schemas. If a precondition of the procedural schema, or its steps, matches the state, RW is verified.

Not all continuation sentences are of the dominant type. When the test for the dominant type fails, all other viable continuation types and coherence relations must be tested in order of dominance. If no relation is found between the sentences, the sentences are interpreted as being unrelated.

We illustrate the processing for finding continuation types with several examples. We begin with two sentence pairs that are similar to our first examples in that they differ only in their quantifiers. However, this difference leads to a different understanding of the relation between the two sentences and to different inferences about the AUVs in addition to different interpretations of the pronoun.

E Few of the AUVs were launched from Tampa.
F They collected samples in Boston Harbor.

G Many of the AUVs were launched from Tampa.
H They collected samples in Boston Harbor.

In both pairs, the first sentence describes a specific instance of the P-LAUNCH schema. That schema is a step in other schemas, including P-MISSION and P-COLLECT-MISSION, which is a specialization of P-MISSION for collecting samples. The second sentence describes a specific instance of P-COLLECT. This schema is the work step in P-COLLECT-MISSION and is executed after P-LAUNCH. Consequently, P-COLLECT-MISSION is found as the schema which links the two steps. This schema also includes a MOVE-TO action which causes the AUV to travel from the launch site to the collection site. Although both continuation type tests begin with the same processing to find the encompassing schema, they differ when testing the preconditions for evidence of the continuation type.

Sentence 3.2.1 contains the quantifier 'few' for which RWN is the dominant continuation type. The test for RWN begins by finding the encompassing schema as described above. There are several preconditions which must be met in order for the second event to successfully follow the first. For example, the AUV must move

[3] The NLPTool interprets actions that are simply used to change states as the state that the action produces. In other systems, and to extend the testing in the NLPTool, results of an action may also be seen as a state which can create the RW relation.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

to the location where the samples will be collected. This means that there must be a waterway, navigable by the AUV, linking the two locations. The AUVs must also have the resources required to go from the launch site to the collection site. Tampa and Boston Harbor are linked by navigable waterways. This connection involves no default information, so no inference can be made to support RWN. However, most AUVs do not have the resources to make the trip from Tampa to Boston Harbor. The default resources of the AUV would not satisfy this precondition. Because the system is testing for RWN, it seeks to show that a precondition will be violated, so no attempt is made to supply a way to satisfy this precondition. Using this knowledge, RWN is verified.

The next sentence pair, Sentences 3.2.1 and 3.2.1, contains the quantifier 'many' for which WHN is the dominant continuation type. The test for WHN also finds the encompassing P-COLLECT-MISSION as described above. However, because WHN is being tested, extra effort is expended to show that the preconditions can be met. In this case, the system finds that the resource limitations are default information that is not necessarily true for all AUVs. It looks for specific types or instances of AUVs that have enough resources to complete the mission. It finds *long-range AUVs*, a class of AUVs that have the additional resources required to travel long distances and survive in the water for extended periods of time. To verify the WHN continuation type, the inference is made that the AUVs are long-range AUVs and the refset is updated to include this information.

Not all continuation types can be verified for all sentence pairs. Even with inferences biasing the outcome, there are cases in which a continuation type cannot be verified because no reasonable inferences can be made to satisfy the preconditions. For example, there are no navigable waterways that connect the launch site and collection site in the following sentence pair:

I Many of the AUVs were launched from Tampa.
J * They collected samples at Lake Winnipesaukee.

When the precondition of connecting waterways is tested, it fails. Since there are no reasonable inferences which can be made to overcome this failure, the WHN test fails. For this sentence pair, the secondary continuation type, RW, also fails. Since the only additional relation suggested by the discourse structure is that the sentences form a narrative and are related by WHN, no relation can be found between the sentences. Although both sentences are understood, the sentence pair appears incoherent because no relation can be found between the sentences.

Tests for secondary continuation types do not always fail. For example, the RW continuation type, the secondary type for 'many', is verified for the following sentence pair:

K Many of the AUVs surfaced.
L They needed to refuel.

WHN fails immediately because the continuation sentence does not relate an action. The action of surfacing is a step in the procedural schema for refueling from a mother ship, so the RW continuation type is verified.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

In the case of sentence pairs involving a negative quantifier, it is particularly important to verify the dominant and secondary continuation type since it has a significant impact on how the sentences' pronominal reference is understood.

Consider the following example:

M Few of the AUVs were launched in Coral Gables.
N They collected samples at Tampa.

Here, no preconditions have been violated. The Atlantic and Gulf of Mexico connect the two locations and the distance between them can be traversed by all existing AUVs. Because the RWN relation cannot be found, the system moves to 'few"s secondary continuation type, WHN. Since all preconditions are met, WHN is verified. Because the WHN continuation type, and not the RWN continuation type, is verified, 'they' is understood to refer to the refset instead of the compset.

3.2.2 Disambiguating referents

Once the continuation type is found, it is easy to find the referent of the pronoun, since the continuation type constrains the sets which can serve as the actor for the action or goal in the continuation sentence as follows:[4]

RWN: compset	WHN: refset
RW: refset	CONS: refset/compset

The test for each continuation type begins by checking that the required set is a potential interpretation for the actor. Since it requires the compset as the referent of 'they', RWN will be ruled out as a possible continuation type for positive quantifiers when a pronominal reference is used for the actor. When the actor is referred to with 'they' and a single set must serve as the actor of the verified continuation type, that set will be the referent for 'they'. We expect the reasoning needed to verify the CONS continuation type to disambiguate the referent. So, when the proper continuation type is found, the pronominal reference is disambiguated.

We return to our first two examples to illustrate how the continuation type helps to identify the referent of 'they'.

E Few of the AUVs were launched from Tampa.
F They collected samples in Boston Harbor.

The RWN continuation type is found and the pronoun is disambiguated by selecting the compset.

G Many of the AUVs were launched in Tampa.
H They collected samples in Boston Harbor.

[4] The compset can serve as the referent in a WHN continuation sentence under certain, well-specified conditions. Our implementation currently does not include these cases. However, because the conditions are well-specified, we could extend our implementation to include these cases by testing for the condition before selecting the referent.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

Although the compset is not in focus, we cannot simply assume the referent of 'they' is the refset because both the refset and the wholeset are possible referents. When the WHN continuation type is verified, the refset is selected as the referent of 'they'.

3.3 Quantifiers and clue words

In our examples, so far, only predictions of coherence relations made by quantifiers are necessary to find the correct interpretations of the pronoun and the correct relations between sentences. However, other sources, such as clue words and discourse structures, also suggest coherence relations. Clue words are explicit markers for the relations between sentences (e.g. Reichman 1985; Cohen 1987; Litman and Allen 1987). As such, they override other sources. However, a clue word cannot force an interpretation of the pronoun that is ruled out by the quantifier.

For example, in the sentence pair below, the clue word 'next' overrides the preferences of the quantifier for the RWN relation and forces the WHN relation.

O Few of the AUVs were launched from Tampa.
P Next, they collected samples in Boston Harbor.

We use the same test for clue words as for quantifiers, so the WHN test proceeds exactly as it did for Sentences 3.2.2 and 3.2.2. Sentences 3.2.2 and 3.2.2 differ from this example in that no clue word is used and the quantifier is 'many', causing WHN to be tested as the dominant continuation type. As for Sentences G and H, the system verifies WHN for Sentences O and P. This causes the system to infer that the AUVs which were launched in Tampa must be long-range AUVs that have the capability to travel to Boston Harbor. The system does not consider the RWN continuation type that is preferred by 'few' and would not consider it even if the WHN test had failed.

When a clue word specifies a relation, the test for that relation can fail for one of two reasons. First, it may fail because the relation cannot be supported by world knowledge. For example, consider the sentence pair:

Q Many of the AUVs were launched from Tampa.
R * They collected samples at Lake Winnipesaukee next.

This sentence pair differs from Sentences 3.2.1 and 3.2.1 only in that a clue word is used to specify the WHN relation. In both cases, the WHN continuation type fails because no route can be found allowing the AUVs to travel from Tampa to Lake Winnipesaukee.

Second, the test can fail because the set needed for the relation is not available. Consider the following pair of sentences:

S Many of the AUVs were launched from Tampa.
T * They collected samples in Boston Harbor instead.

Here, the clue word 'instead' signals the RWN relation. RWN requires the compset as the actor of the action in the continuation sentence. However, the compset is not

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

in focus at the end of the first sentence. When the RWN test checks if the compset is available, that check fails and the RWN test fails without further reasoning about the relation between the sentences.

4 Discussion

We have discussed the difficulty in understanding pronominal references to quantified expressions and have shown how they can be understood in a natural language system. Our approach relies on recognizing the continuation type of the sentence which contains the pronoun. Continuation types are tested in order of dominance, so the preferred interpretation of the pronoun is found in cases in which other possible interpretations may exist.

Most natural language systems, by necessity, already have most of the components needed to correctly handle pronominal references to quantified expressions. These include a mechanism for placing sets in focus and a method for determining the coherence relation between sentences. These systems can be easily extended to handle pronominal references to quantified expressions by taking information associated with the quantifier into consideration. In this paper, we have described these extensions in detail and have presented examples from an extended system.

Our system has been tested on 25 pairs of sentences, including example Sentences E–T in this paper. Three quantifiers (few, many and some) and two clue words (next and instead) appear in the test sentences. Sentence pairs are related by the WHN, RWN or RW continuation type or are not related. Among sentence pairs for which a relation is found, test sentence pairs are related by either the dominant or secondary continuation type. The test suite also includes pairs of sentences which were found to be unacceptable because no relation was found.

Our current system implements a framework for testing the general approach. Our results suggest that the work is worth continuing, but we need to expand several parts of the system to make it more robust. For example, we have implemented fairly simple tests for continuation types that work in a limited domain. Clearly, as we add more world knowledge and more thorough tests, more examples will be handled by the system. However, as this capacity is expanded, we will face an important area of future work: the need to determine how much effort is too much effort when testing a relation. This will be important for determining when to move from the dominant to the secondary continuation type, as well as indicating when clue words should be generated. Another area of future work is the problem of integrating quantifier preferences with preferences for coherence relations from other sources, specifically discourse structure. In addition, we would like to explore how quantifiers affect processing when pronouns do not appear in the continuation sentences and other, lower-focus, referring expressions are seen.

References

Allen, J. (1995) *Natural Language Understanding*. Benjamin/Cummings.

Brennan, S. E., Friedman, M. W. and Pollard, C. J. (1987) A centering approach to pronouns. *Proceedings of 25th Annual Meeting of the ACL*, pp. 155–162. Stanford, CA.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

Cohen, R. (1987) Interpreting clues in conjunction with processing restrictions in arguments and discourse. *Proceedings of the Sixth National Conference on Artificial Intelligence*, pp. 528–533. Morgan Kaufmann.

Corblin, F. (1996) Quantification et anaphore discursive: La référence auc complémentaires. *Languages 123*, pp. 51–74.

Cullingford, R. E. (1986) *Natural Language Processing: A Knowledge Engineering Approach*. Rowan and Littlefield.

Fincher-Kiefer, R. (1996) Encoding differences between bridging and predictive inferences. *Discourse Processes*, **22**(3): 225–246.

Grosz, B. J. (1977) The representation and use of focus in a system for understanding dialogs. *Proceedings of the Fifth International Conference on Artificial Intelligence*, pp. 67–76. William Kaufmann.

Grosz, B. J., Joshi, A. K. and Weinstein, S. (1995) Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, **21**(2): 203–226.

Grosz, B. J. and Sidner, C. L. (1986) Attention, intention, and the structure of discourse. *Computational Linguistics*, **12**(3): 175–204.

Hartley, J., Trueman, M. and Rogers, K. A. (1984) The effects of verbal and numerical quantifiers on questionnaire responses. *Applied Ergonomics*, **11**: 149–155.

Heim, I. (1982) *The Semantics of Definite and Indefinite Noun Phrases*. PhD thesis, University of Massachusetts.

Hobbs, J. R. (1979) Coherence and coreference. *Cognitive Science*, **3**(1): 67–90.

Holyoak, K. J. and Glass, A. L. (1978) Recognition confusions among quantifiers. *Journal of Verbal Learning and Verbal Behavior*, **17**: 249–264.

Kamp, H. (1981) A theory of truth and semantic representation. In: Groendijk, J., Janssen, T. and Stochof, M., editors, *Formal Methods in the Study of Language*, pp. 277–321. Mathematical Centre, Amsterdam.

Kamp, H. and Reyle, U. (1993) *From Discourse to Logic*. Kluwer.

Kibble, R. (1997) Complement anaphora, monotonicity and dynamic binding. In: Ploch, S. and Swinburne, D., editors, *SOAS Working Papers in Linguistics and Phonetics 7*. Department of Linguistics, School of Oriental and African Studies.

Lappin, S. and Leass, H. J. (1994) An algorithm for pronominal anaphora resolution. *Computational Linguistics*, **20**(4): 535–561.

Litman, D. J. and Allen, J. F. (1987) A plan recognition model for subdialogues in conversation. *Cognitive Science*, **11**: 163–200.

Mann, W. C., Moore, J. A. and Levin, J. A. (1977) A comprehension model for human dialogue. *Proceedings of the Fifth International Conference on Artificial Intelligence*, pp. 77–87. William Kaufmann.

Matthias, C. M. (1997) Understanding references to quantified expressions. Technical Report 97-03, University of New Hampshire. Master's thesis.

McKeown, K. R. (1985) *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge University Press.

Moxey, L. M. and Sanford, A. J. (1993a) *Communicating Quantities: A Psychological Perspective*. Lawrence Erlbaum.

Moxey, L. M. and Sanford, A. J. (1993b) Prior expectation and the interpretation of natural language quantifiers. *Euro. J. Cognitive Psychology*, **5**(1): 73–91.

Reichman, R. (1985) *Getting Computers to Talk Like You and Me: Discourse Context, Focus, and Semantics (An ATN Model)*. MIT Press.

Sanford, A. J., Moxey, L. M. and Paterson, K. B. (1996) Attentional focus with quantifiers in production and comprehension. *Memory and Cognition*, **24**(2): 144–155.

Sidner, C. L. (1983) Focusing in the comprehension of definite anaphora. In: Brady, M. and Berwick, R., editors, *Computational Models of Discourse*. MIT Press.

Turner, R. M. (1994) *Adaptive Reasoning for Real-World Problems: A Schema-Based Approach*. Lawrence Erlbaum.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press

Turner, R. M., Blidberg, D. R., Fox, J. S. and Turner, E. H. (1991) Multiple autonomous vehicle imaging system (MAVIS). *Proceedings of the 7th International Symposium on Unmanned Untethered Underwater Submersible Technology (AUV '91)*.

Wilensky, R. (1978) Understanding goal-based stories. Technical Report 140, Department of Computer Science, Yale University. PhD thesis.

Wright, D. B., Gaskell, G. D. and O'Muircheartaigh, C. A. (1994) How much is 'Quite a bit'? Mapping between numerical values and vague quantifiers. *Applied Cognitive Psychology*, **8**: 479–496.

Zadeh, L. A. (1965) Fuzzy sets. *Information and Control*, **8**: 338–353.

https://doi.org/10.1017/S1351324999002089 Published online by Cambridge University Press