A Centering Analysis of Relative Clauses in English and Greek

Eleni Miltsakaki

1 Introduction

Centering Theory (Grosz et al., 1995) was developed as a model of local coherence in discourse. Coherence in Centering is evaluated in terms of center transitions, defined and ranked to reflect four degrees of coherence. Center transitions are computed for each 'utterance' by means of two basic comparisons; whether the topic of the current utterance is the same as the topic of the preceding utterance and whether it is realized in a prominent position, e.g., in subject position in English. An entity realized in subject position is the highest ranked entity of the utterance and most likely center of the succeeding discourse. Other entities in the same utterance are ranked lower.

Complex sentences raise interesting questions for Centering as well as our understanding of the effects of syntactic stucture on discourse processing. This is because complex sentences contain multiple subjects, i.e., the subject of the main clause and the subjects of subordinate clauses. What the salience status of these entities is with respect to each other is a question that merits further investigation. Following up on earlier work on the salience status of entities in adverbial clauses (Miltsakaki, 2002a,b), here we focus on relative clauses. We compare and contrast the salience of entities in main and relative clauses in two languages: English and Greek.

For each language 100 tokens of non-restrictive, sentence final relative clauses were extracted from corpora. Centering transitions were computed in two conditions: the complex sentence condition, in which the relative clause was processed as a single unit with the main clause, and the single clause condition, in which the main and relative clause were processed as a single unit. Assuming reasonable coherence in the written corpus, we expected that if entities evoked in relative clauses were of equal salience status as entities evoked in main clauses, then the single clause condition should yield more 'coherent' transitions than the complex sentence condition. Preliminary analysis of the results shows that this is not the case. Most importantly, closer inspection of choice and interpretation of referential forms in the following discourse shows that subjects of relative clauses do not always warrant pronominal reference even in the absence of more recent competing antecedents.

The remaining of the paper is organized as follows. Section 2 summarizes

previous work on entity salience in complex sentences. In Section 3, we give a brief overview of Centering Theory. The Centering studies on English and Greek are presented in Sections 4 and 5 respectively. We conclude with a few thoughts on the implications of the current results with respect to the nature of the interaction between topics, subjects and pronouns.

2 Previous Work on Complex Sentences

Kameyama (1998, 1993) suggests breaking up complex sentences according to the following hypotheses: (i) Conjoined and adjoined tensed clauses form independent center update units. (ii) Tenseless subordinate clauses, report complements and relative clauses belong to the update unit containing the main clause. With regard to the tensed adjunct hypothesis, which treats tensed adjunct clauses as independent units, Kameyama (1998) brings support from backward anaphora. She argues that the tensed adjunct hypothesis predicts that a pronoun in a preposed subordinate clause is anaphorically dependent on an entity already introduced in the immediate discourse and not on the subject of the main clause it is attached to. However, this argument is challenged by empirical data. Carden (1982), van Hoek (1997), and Tanaka (2000) provide empirical evidence of pronouns which are the first mention of their referent in discourse.

In (Miltsakaki, 2002a), the interpretation of pronominal subjects was tested in two conditions in a sentence completion task. In the first condition, main-main condition, the subject pronoun was in the second main clause. In the second condition, main-subordinate condition, the subject pronoun was in the second adverbial clause. Sample critical items of the main-main and main-subordinate conditions are shown in (1) to (4).

(1)   The groom hit the best man. Moreover, he...

(2)   The beggar pushed the gentleman so that he...

(3)   The boxer kicked the referee. As a result, he...

(4)   The policeman shot the burglar because he...

The results of this study showed a strong main effect of the type of the clause type ($F(1,19)=79.33$, $p<0.001$). Specifically, in the main-subordinate condition, the interpretation of the subject pronoun varied between the subject and the object of the main clause. In contrast, in the main-main condition, the subject pronoun showed a very strong tendency to be interpreted as the subject of the main clause. Taken together these results indicate that pronominal interpretation is sensitive to the distinction between main and subordinate

clauses. Across main clauses, subjects are more salient than objects and their referents are consistently picked for the interpretation of the subsequent subject pronoun. However, when a subject pronoun is in a subordinate clause, its interpretation varies. We conclude that subject salience is strong across main clauses but intra-sententially subject salience may not be the primary factor for pronominal interpretation.

These results are consistent with the hypothesis that, in Centering, topics are updated across sentences, i.e., that main and subordinate clauses are processed as one unit. In the main-main condition, the highest ranked entity in the first main clause is the subject, and most likely topic of the next sentence. The interpretation of the succeeding pronoun is then correctly predicted to be the subject of the first main clause. If, indeed, a main and subordinate clause form a single unit of topic update, the next question to be addressed concerns the salience status of entities evoked in a subordinate clause after the unit is processed. This question is equivalent to asking what determines the relative salience of entities within the complex structure.

Cooreman and Sanford (1996) investigated the interpretation of a subject pronoun following a main and an adverbial clause, each introducing a same gender referent. In a sentence completion task, they presented participants with a complex sentence containing a main and an adverbial clause. Participants were prompted to start a continuation with a pronoun, which could be interpreted either as the entity introduced in the main clause or the entity introduced in the adverbial clause. To check for clause order effects, the adverbial clause appeared both after and before the main clause. Three sets of subordinate conjunctions were used: 'after/before', 'when/while', and 'because/since'. Sample items are shown in examples (5a) and (5b).

(5)   a. After the tenor opened his music store the conductor sneezed three times.
            He...

        b. The conductor sneezed three times after the tenor opened his music score.
            He...

The results of this experiment revealed that for all three sets of connectors the main clause referent was the preferred choice for the interpretation of the pronoun in the continuation: 92.9% for 'after/before', 80.3% for 'when/while', and 79.8% for 'because/since'. The order in which the main and adverbial clauses were presented did not make a difference except for the subordinate conjunction 'because': the main clause referent was the preferred choice for the interpretation of the pronoun in the continuation 75.2% in the main-subordinate order versus 85.4% in the subordinate-main order. Cooreman and Sanford (1996) report that there was no such effect for any other

subordinate conjunction, including 'since'.

3 Overview of Centering Theory

Centering theory was developed as a model of the relationship between discourse coherence, discourse structure and choice of referring expression (Grosz et al., 1995). What we perceive as the topic of an utterance, at least in the sense of Reinhart (1981); Horn (1986), among others, is formally defined as the Backward-looking center. Each utterance evokes a list of *Forward-looking centers*, ranked according to degree of salience. The highest ranked entity on the list of Forward-looking centers is called the *Preferred center*. The *Backward-looking center* is the highest ranked entity of the preceding utterance that is realized in the current utterance.

The Centering model is designed to capture those aspects of processing that are responsible for the difference in the perceived coherence of discourses as those demonstrated in (6) and (7) below (Grosz et al., 1995).

(6)   a.   John went to his favorite music store to buy a piano.

       b.   He had frequented the store for many years.

       c.   He was excited that he could finally buy a piano.

       d.   He arrived just as the store was closing for the day.

(7)   a.   John went to his favorite music store to buy a piano.

       b.   It was a store John had frequented for many years.

       c.   He was excited that he could finally buy a piano.

       d.   It was closing just as John arrived.

Discourse (6) is intuitively more coherent than discourse (7). This difference may be seen to arise from the different degrees of continuity in what the discourse is about. Discourse (6) centers a single individual, 'John', whereas discourse (7) seems to center in and out on different entities, 'John', 'store', 'John', 'store'. Degrees of continuity are reflected in four *Centering transitions*. These are: *Continue, Retain, Smooth-Shift* and *Rough-Shift*. Centering transitions are computed according to Table 1. In Table 1, the Backward-looking center is designated as Cb and the Preferred center as Cp. The current utterance is shown as U_{i-1} and the preceding utterance as U_{i-1}. The most coherent transition is a Continue, identified when the topic of the current utterance, $Cb(U_i)$ is the same as the topic of the previous utterance, $Cb(U_{i-1})$, and is realized in a prominent position (Cp), e.g., in subject position. The

	$Cb(U_i)=Cb(U_{i-1})$	$Cb(U_i)\neq Cb(U_{i-1})$
$Cb(U_i)=Cp$	Continue	Smooth-shift
$Cb(U_i)\neq Cp$	Retain	Rough-shift

Table 1: Table of Centering transitions

least coherent transition is a Rough-Shift, identified when the topic of the current utterance, $Cb(U_i)$, is not the same as the topic of the preceding utterance, $Cb(U_{i-1})$, and it is not realized in a prominent position (Cp).

Interpreting a pronominal reference as the Backward-looking center in an utterance is captured in Centering's *Pronoun rule* which predicts that if there is a single pronoun in an utterance, then this pronoun refers to the current topic. Of course, the Pronoun rule holds in utterances with more that one pronoun but the rule makes a prediction for only one of these pronouns. So, in Centering, subjects, topics and pronouns are related via the notions of Preferred center and Backward-looking center.

For the computation of topic transitions as well as the empirical evaluation of the Pronoun rule, the definition of the *utterance*, i.e., the topic update unit, is crucial. In the original formulation of Centering, the question of the extent of the utterance was left open to empirical investigation. Here, I will sketch out the predictions that will be tested to determine whether subordinate clauses are processed as an utterance. If each tensed clause, main or subordinate, determines the extent of an utterance, we expect that a succeeding pronoun, whether in a main or subordinate clause, will co-specify with the current topic, which in experimental conditions can be expressed by the subject of the preceding clause. In a corpus, the topic of an utterance cannot be identified independently. The extent of an utterance will be tested by comparing Centering transitions in two conditions: processing each tensed clause as a unit and processing the complex sentence as a unit. The condition yielding more coherent transitions will be taken to reflect the appropriate extent of an utterance, assuming that written text is maximally coherent. Prior Centering analyses of corpora have indeed shown that Rough-Shift transitions, for example, do not occur in written text (Di Eugenio, 1998, among others).

For the ranking of entities in the list of Forward-looking centers, entities are ranked according to grammatical role as suggested in Brennan et al. (1987); Walker et al. (1998), among others. Subjects rank higher than objects which rank higher than other entities. For reasons discussed in detail in Reinhart (1981) quantificational expressions, non-specific indefinite phrases as well as impersonal references (Prince, 1999b) are either not included in the list of Forward-looking centers or ranked low.

4 Centering Analysis of English Relative Clauses

For the Centering analysis of English relative clauses, 100 tokens of non-restrictive relative clauses were extracted from the Wall Street Journal corpus. Extraction was done according to the following criteria: a) the relative clause was preceded by a comma (to exclude restrictive relative clauses), b) the sentence following the relative clause included reference to at least one entity evoked in the sentence containing the relative clause, either in the main clause or in the relative clause, and c) the relative clause was in sentence-final position (to ensure that the relative clause is adjacent to the following unit).

For each token, Centering transitions were computed in two conditions. In condition A, two Centering transitions were computed: one for the sentence containing the relative clause and one for the sentence following the relative clause. In other words in condition A, the center update unit is the complex sentence. Let's call this the *complex sentence condition*. In condition B, three Centering transitions were computed. One for the first sentence excluding the relative clause, one for the relative clause and one for the sentence following the relative clause. Condition B, then, assumes that each clause, either main or relative is an independent unit. Let's call it the *single clause condition*.

The resulting Centering transitions in the two conditions are shown in Table 2. Specifically, Table 2 shows the results for the single clause version. The column "more coherent transition" contains the number of cases where a more 'coherent' transition was computed in the final sentence. The column "less coherent transition" shows how many times a less 'coherent' transition was computed in the final sentence, and, finally, the column "no effect" shows how many times the same transition was computed in both conditions. The relevant degree of coherence was specified according to the Centering transitions rule: Continue > Retain > Smooth-Shift > Rough-Shift. So, for example, if the transition computed for the unit following the relative clause was Continue in the single clause condition but Rough-Shift in the complex sentence condition, then the transition was identified in Table 2 as more 'coherent'. Reversely, if the transition computed for the last unit is, for example, Smooth-Shift in the single clause condition and Continue in the complex sentence condition, then the transition was identified as less 'coherent'.

More coherent transition	Less coherent transition	No effect	Total
13	46	41	100

Table 2: Effect of non-restrictive relatives on Centering transitions

A typical example of the category "less coherent transition" is given in (8) and (9). The computation of transitions in the single clause condition shown in (8) yields a Rough-Shift transition, which is classified as less 'coherent' than the Continue transition computed in the complex sentence transition, shown in (9).

(8)   (A disaffected, hard-drinking, nearly-30 **hero**$_i$ sets off for snow country in search of an elusive sheep with a star on its back at the behest of a sinister, erudite mobster with a Stanford degree.)
        SINGLE CLAUSE CONDITION

    a. **He**$_i$ has in tow his prescient **girlfriend**$_j$,
        Cb=hero Cp=hero Tr=Continue

    b. **whose**$_j$ sassy retorts mark her as anything but a docile butterfly.
        Cb=girlfriend Cp=girlfriend Tr=Smooth-Shift

    c. Along the way, **he**$_i$ meets a solicitous Christian chauffeur who offers the hero God's phone number;
        Cb=none Cp=hero Tr=Rough-Shift

(9)   COMPLEX SENTENCE CONDITION

    a. **He**$_i$ has in tow his prescient girlfriend, whose sassy retorts mark her as anything but a docile butterfly.
        Cb=hero Cp=hero Tr=Continue

    b. Along the way, **he**$_i$ meets a solicitous Christian chauffeur who offers the hero God's phone number;
        Cb=hero Cp=hero Tr=Continue

Examples such as the above are supportive of the hypothesis that the relative clause is not processed as topic update unit. Processing the relative clause as a unit by itself yields three problems. First, we would process the 'girlfriend' as the most likely topic of the subsequent discourse, an expectation that is not met. In fact, this entity is not mentioned at all in the following sentence. Second, counter to intutition, the discourse would be modeled as disconnected. Disconnected discourses are predicted to be hard to process because they place on the hearer the extra burden of inferring the intended link. Third, the use of the pronoun would be puzzling. If the most salient entity after processing the relative clause is the 'girlfriend' then the pronominalized reference to the 'hero' which was evoked two units before is unexpected.

On the other hand, if the discourse is processed according to the complex sentence hypothesis, then none of the problems above arises. The highest ranked entity in that unit, the 'hero', is processed as the most likely topic of the discourse, an expectation that is met as indicated by the pronominal reference

in the subject position. The discourse now 'coheres' in that it is perceived as being about the same entity.

Returning to the results of this study in Table 2, the single clause condition yields a more 'coherent' transition in 13 of the 100 cases, whereas for 41 cases, it yields a less 'coherent' transition. So, overall these findings lend support to the complex sentence hypothesis. But what about the 13 instances in which the single clause hypothesis appears to yield more 'coherent' transitions?

The example shown in (10) and (11) is representative of the cases in which the single clause condition yields a more 'coherent' transition. In this case, processing the relative clause as an independent unit yields a Continue transition, which is more coherent than the Smooth-Shift transition computed in the complex sentence condition. Closer inspection of this example, however, reveals that the head noun is referenced in the subsequent discourse with a full NP, despite the fact that it appears in subject position in the relative clause. This pattern of reference in which an entity is promoted to a subject position with an NP form, has been observed in other languages (Miltsakaki, 2003; Turan, 1995) as a strategy used by speakers to signal a shift to a *new* topic. An entity first evoked in a non-salient position is then promoted to a subject position with a full NP and is established as the new topic. Pronominal reference is avoided in this case, despite the accessibility of the referent, presumably because the referent was not the topic of the sentence in which it was evoked but is intended to be the new topic. Other independent factors that could account for the use of an NP do not hold here. Specifically, the use of the NP in this case is not dictated by the grammar, does not provide any further information about the referent (Fox, 1987) and it does not appear on a segment boundary (Passonneau and Litman, 1993). In any case, other factors licensing the use of an NP perform functions that are independent of referent accessibility, so we should still be able to use a pronoun to refer to 'Mr. Kilpatrick' successfully. This is not the case. According to native speakers' judgment, the preferred interpretation for a subject pronoun in the last sentence would be 'Wilson Taylor', the main clause subject. It seems then that in this example, we do, in fact, have a Smooth-Shift transition which correctly reflects processing of 'Mr. Kilpatrick' as the new topic.

(10)   SINGLE CLAUSE CONDITION

   a. **Wilson H. Taylor**$_i$, president and chief executive officer of this insurance and financial services concern, was elected to the additional post of chairman.

   b. **Mr. Taylor**$_i$, 45 years old, succeeds **Robert D. Kilpatrick**$_j$, 64,
       Cb=Taylor Cp=Taylor Tr=Continue

    c.   **who**$_j$ is retiring, as reported earlier.
        Cb=Kilpatrick Cp=Kilpatrick Tr=Smooth-Shift

    d.   **Mr. Kilpatrick**$_j$ will remain a director.
        Cb=Kilpatrick Cp=Kilpatrick Tr=Continue

(11)   COMPLEX SENTENCE CONDITION

    a.   **Wilson H. Taylor**$_i$, president and chief executive officer of this insurance and financial services concern, was elected to the additional post of chairman.

    b.   **Mr. Taylor**$_i$, 45 years old, succeeds **Robert D. Kilpatrick**$_j$, 64, who is retiring, as reported earlier.
        Cb=Taylor Cp=Taylor Tr=Continue

    c.   **Mr. Kilpatrick**$_j$ will remain a director.
        Cb=Kilpatrick Cp=Kilpatrick Tr=Smooth-Shift

5 Centering Analysis of Greek Relative Clauses

In Greek, relative clauses can be introduced either by the relative pronoun *o opios* or by the complementizer *pu* (null complementizers are not allowed). The expression *o opios* must agree in gender and number with the noun phrase it modifies and it must be in the case appropriate to its grammatical role in the relative clause. It is also possible for the relative pronoun to be the noun phrase complement of a prepositional phrase. In such cases the preposition is always followed by the *o opios*-paradigm. A preposition cannot combine with *pu*.

    In Greek, as in English, relative clauses can be restrictive or non-restrictive. For the Centering analyis of Greek relative clauses, we extracted 100 tokens of non-restrictive, sentence final relative clauses. The corpus consisted mainly of newspaper articles, available on line at http://www.enet.gr. The selection criteria and the computations of Centering transitions were the same as in the English study. The results of the computation of Centering transitions in the two conditions are shown in Table (3).

More coherent transition	Less coherent transition	No effect	Total
8	44	48	100

Table 3: Effect of Greek non-restrictive relatives on Centering transitions

    For the Greek data, the examples are glossed and translated in English and then the two conditions are presented with the English translations. In the case

of examples (12) and (13), condition B yields a Rough-Shift transition, which is ranked less 'coherent' than the Continue transition computed in condition A for the same last sentence. What is important to note in this example is that in condition B, where the relative clause is processed as an independent unit, there is no link between the relative clause and the subsequent utterance. Note that in the last sentence, *Kostas Karamanlis* has been coded as the Cp although he is not the subject. We did so because the subject is a negated indefinite NP. As mentioned earlier, a possibility that we have not fully investigated is that *Kostas Karamanlis* outranks the subject due to animacy. However, even if we code the subject as the Cp, we still get a less coherent transition in condition B.

(12) a. I   prothesi tu Kosta Karamanli$_i$ itan safestati   htes   sti
          The intention of Kosta Karamanli was most-clear yesterday at-the
          sindedriasi$_j$ tis   ektelestikis epitropis tu   komatos, apo tin opia$_j$
          meeting   of-the executive   committee of-the party,   from the which
          apusiazan   i   Mil. Evert ke   Ntora Mpakogiani.
          were-missing the Mil. Evert and Ntora Mpakogiani.

          '**Kosta Karamanli's**$_i$ intention was very clear yesterday at **the meeting**$_j$ of the party's executive committee, from **which**$_j$ (meeting) Mil. Evert and Ntora Mpakogiani were missing.'

      b. Omos   apo tin topothetisi tu   proedru$_i$ tis   ND den apusiase
          However from the position   of-the president of-the ND not was-missing
          ki   enos   idus ehmi gia to dimarho tis   Athinas ke   to endehomeno
          and of-one kind dig   for the mayor   of-the Athens and the possibility
          dimiurgias   neu   komatos apo afton.
          of-founding of-new party   by   HIM

          'However, in the position of **ND's president**$_i$ a kind of dig was present against the mayor of Athens and the possibility of the founding of a new party by HIM.

(13) CONDITION A

      a. **Kosta Karamanli's**$_i$ intention was very clear yesterday at **the meeting**$_j$ of the party's executive committee, from **which**$_j$ (meeting) Mil. Evert and Ntora Mpakogiani were missing.
          Cb=none, Cp=Karamanlis, Tr=none

      b. However, in the position of **ND's president**$_i$ a kind of dig was present against the mayor of Athens and the possibility of the founding of a new party by HIM.
          Cb=Karamanlis, Cp=Karamanlis, Tr=Continue

(14)   CONDITION B

    a. **Kosta Karamanli's**$_i$ intention was very clear yesterday at **the meeting**$_j$ of the party's executive committee,
Cb=none, Cp=Karamanlis, Tr=None

    b. from **which**$_j$ (meeting) Mil. Evert and Ntora Mpakogiani were missing.
Cb=meeting, Cp=Evert and Mpakogiani, Tr=Retain

    c. However, in the position of **ND's president**$_i$ a kind of dig was present against the mayor of Athens and the possibility of the founding of a new party by HIM.
Cb=none, Cp=Karamanlis (ND's president), Tr=Rough-Shift

Looking back at Table 3, we see that for 44% of the tokens, processing the relative clause as an independent unit yields a less 'coherent' Centering transition in the subsequent sentence, whereas a more 'coherent' transition was computed for only 8% of the tokens. For the remaining 48%, there was no effect. If we now look closer at the 8 tokens of the more 'coherent' transition we observe that our coding schema actually failed to capture implicit links present in the relevant discourses. This was the case for 7 of the 8 examples. In (15), for example, *British documents* and *British Ambassador* are linked via functional dependency. However, this link is missed in the computation of Centering transitions because the two do not co-refer.

(15)   Sta   vretanika egrafa$_i$ tu   Iuliu 1970 anaferete   oti o
In-the british   documents of-the July 1970 is-mentioned that the
Papadopulos ihe plirofories   gia dolofoniki apopira enantion tu   Pipineli.
Papadopulos had information for murder   attempt against   of-the Pipineli

'In the **British**$_i$ documents of July 1970 it is mentioned that Papadopulos had information about a murder attempt against Pipineli.'

    a. O   Vretanos presveftis$_j$ epikalite 'to gnosto kikloma sizigon',   ...,
The British   ambassador appeals 'the known circle   of-wives$_{kl}$', ...,
me   tis opies$_k$ o   ser Maikl   Stiuart$_j$ diatiruse   filikes
with the whom$_{kl}$ the sir Michael Steward maintained friendly
shesis.
relationships

'**The British Ambassador**$_j$ appeals to 'the well-known **circle of wives**$_{kl}$', ..., with **whom**$_{kl}$ Sir Michael Steward had friendly relationships.'

    b. I   sizigos$_l$ tu   Pipineli diohetefse tin pliroforia   afti ston   Stiuart
the wife   of-the Pipineli released   the information this to-the Steward
ke   ...
and ...

'**Pipineli's wife**$_l$ released this information to **Steward**$_j$ and ... '

6 Discussion and Conclusions

The design of the corpus studies on relative clauses was based on basic principles of Centering theory. However, the significance of the findings goes beyond the Centering framework. Centering formalizes basic intuitions that we have about entity based coherence. Discourses that are carefully planned around a single entity and smoothly shift our attention to new topics are perceived more coherent than discourses that either focus entities in and out or discourses that appear to be disconnected. Naturally, when talking about topics and topic continuity we need to address the issue of the topic update unit. Although topic update units may or may not turn out to be defined on structural grounds as suggested here, the Centering studies in English and Greek have revealed that there is a possibly cross-linguistic distinction between main and subordinate clauses which challenges the tacit assumption that all clauses are born equal: topic identification, entity salience and choice of referring expression appear to be sensitive to the syntactic choices made by the speakers when they organize discourse.

These observations raise new questions that need to be addressed. First, it is not clear what property of subordinate clauses is responsible for the attested patterns. Subordinate clauses are introduced with subordinate conjunctions which express a discourse relation between the main and the subordinate clause. Adverbial clauses express a variety of relations, e.g., causality, concession, temporal sequence etc. Non-restrictive relative clauses can typically be interpreted as elaborations on the discourse entity evoked with the head noun. Recently, Kehler (2002) proposed a promising account on the relationship between coherence relations and reference. In the proposed framework pronouns are treated as variables whose interpretation is contingent on the type of coherence relation established in the on-going discourse and falls out of the semantic representation. Three basic relations are identified: *resemblance*, *cause-effect*, and *contiguity*. Skipping the details of the proposed theory, the establishment of a resemblance relation basically accounts for data which show subjects to be the preferred antecedents. A cause-effect relation accounts for data which may show an object to be the most preferred antecedent. Interestingly, representative connectives of resemblance relations belong to the class of sentence adverbials, such as 'however' and 'for example'. Representative connectives of cause-effect relations are subordinate conjunctions, such as 'because' and 'even though', with the exception of 'and as a result' and 'and therefore'. It would be useful to see if the effects of the main-subordinate distinction can be uniformly attributed to coherence relations as defined in Kehler (2002). Relative clauses, though, would still require special investigation, as in the current

formulation of Kehler's theory they remain unexplored territory.

The main-subordinate distinction raises further questions on the nature of subordination itself. Syntactic subordination of the type we are concerned with here seems to be a universal property of languages. Still, little is understood as to why subordinate clauses exist in grammar. With the exception of complement clauses and restrictive relative clauses, subordinate clauses such as adverbial and non-restrictive relative clauses do not participate in the predicate-argument structure of the verb. As a toy experiment one could successfully rewrite any text as a succession of main clauses without changing propositional content or discourse relations (e.g., (16) and (17)). While the investigation of the nature and purpose of subordination in grammar will await further research, the fact remains that grammars allow speakers to choose between using a subordinate or non-subordinate clause to express propositions and relations between them.

(16)   a.   Mary was late this morning because she missed the 8am bus.

       b.   Mary missed the 8am bus this morning. As a result, she was late.

       c.   Mary was late this morning. She missed the 8am bus.

(17)   a.   Mary was hired by Mr Brown, who is the director of NBC.

       b.   Mary was hired by Mr Brown. Mr Brown is the director of NBC.

Prior study on the relationship between choice of linguistic form and discourse function has shown that speakers use syntactic variability to express a variety of discourse functions which contribute a range of meanings that are not derived compositionally from the syntactic representation (Prince, 1999a; Ward and Birner, 1996; Vallduvi and Vilkuna, 1998, among others). It is possible that *one* of the factors driving the choice of subordinate clauses in discourse might be to mark entities with low salience. Strategies of marking low salience can reduce the complexity of inferencing required in processing discourse. Joshi and Kuhn (1979) proposed a centered logic approach to discourse processing according to which one of the entities in the processed discourse is singled out to form a special argument, the discourse center. A complex predicate is then constructed, including other entities, which is predicated of the centered entity. In such a model, subordinate clauses can be seen as delimiting the boundaries of the internal structure that will be temporarily hidden in the complex predicate. In this way, it is possible to retain a single center while introducing multiple entities that are possibly propositionally related to the center.

The research studies discussed in this paper raise more questions than it answers. A successful theory that will account for the complexities of refer-

ence and salience in discourse is still to come. However, I hope that the current findings from the study of complex sentences will, in fact, spur new research interest in complex structure and the study of subordination in discourse.

References

Brennan, Susan, Marilyn Walker-Friedman, and Carl Pollard. 1987. A Centering approach to pronouns. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, 155–162. Stanford, CA.

Carden, Guy. 1982. Backwards anaphora in discourse context. *Journal of Linguistics* 18:361–387.

Cooreman, Ann, and Anthony Sanford. 1996. Focus and syntactic subordination in discourse. Technical report, Human Communication Research Center.

Di Eugenio, Barbara. 1998. Centering in Italian. In *Centering theory in discourse*, ed. M. Walker, A. Joshi, and E. Prince, 115–137. Clarendon Press: Oxford.

Fox, Barbara. 1987. *Discourse structure and anaphora*. Cambridge: Cambridge University Press.

Grosz, Barbara, Aravind Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling local coherence in discourse. *Computational Linguistics* 21:203–225.

van Hoek, Karen. 1997. *Anaphora and conceptual structure*. Chicago: University of Chicago Press.

Horn, Laurence. 1986. Presupposition, theme and variations. In *22nd Annual Meeting of the Chicago Linguistics Society*, volume 22, 168–192. Chicago.

Joshi, Aravind, and Steven. Kuhn. 1979. Centered logic: The role of entity centered sentence representation in natural language inferencing. In *6th International Joint Conference on Artificial Intelligence*, 435–439. Tokyo.

Kameyama, Megumi. 1993. Intrasentential Centering. In *Proceedings of the Workshop on Centering*. University of Pennsylvania.

Kameyama, Megumi. 1998. Intrasentential Centering: A case study. In *Centering theory in discourse*, ed. M. Walker, A. Joshi, and E. Prince, 89–112. Clarendon Press: Oxford.

Kehler, Andrew. 2002. *Coherence, reference, and the theory of grammar*. CSLI Publications, Stanford.

Miltsakaki, Eleni. 2001. On the interpretation of weak and strong pronominals in Greek. In *Proceedings of the 5th International Conference in Greek Linguistics*. Sorbonne.

Miltsakaki, Eleni. 2002a. Effects of subordination on referential form and interpretation. In *Proceedings of the 26th Penn Linguistics Colloquium.* Penn Working Papers in Linguistics, Philadelphia.

Miltsakaki, Eleni. 2002b. Towards an aposynthesis of topic continuity and intra-sentential anaphora. *Computational Linguistics* 28:319–355.

Miltsakaki, Eleni. 2003. The syntax-discourse interface: Effects of the main-subordinate distinction on attention structure. Doctoral Dissertation, University of Pennsylvania.

Passonneau, Rebecca, and Diane Litman. 1993. Intention-based segmentation: Human reliability and correlation with linguistic cues. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, volume 19, 3, 148–155. Columbus, OH.

Prince, Ellen. 1999a. *How not to mark topics: 'topicalization' in English and Yiddish*, chapter 8. Austin: University of Texas.

Prince, Ellen. 1999b. Subject pro-drop in yiddish. In *Focus: Linguistic, cognitive, and computational perspectives*, ed. Peter Bosch and Rob van der Sandt, 82–101. Cambrige University Press.

Reinhart, Tanya. 1981. Pragmatics and linguistics: An analysis of sentence topics. *Philosophica* 27:53–93.

Tanaka, Izumi. 2000. Cataphoric personal pronouns in English news reportage. In *Proceedings of the 3rd International Conference of Discourse Anaphora and Anaphor Resolution Conference*, volume 12, 108–117. DAARC 2000, Lancaster.

Turan, Umit. 1995. Null vs. overt subjects in turkish discourse: A Centering analysis. Doctoral Dissertation, University of Pennsylvania.

Vallduvi, Enric, and Maria Vilkuna. 1998. On rheme and kontrast. *Syntax and Semantics* 29.

Walker, M., A. Joshi, and E. Prince, ed. 1998. *Centering theory in discourse*. Clarendon Press: Oxford.

Ward, Gregory, and Betty Birner. 1996. On the discourse function of rightward movement in english. In *Conceptual structure, discourse and language*, ed. Adele Goldberg, 463–479. Stanford: Center for the Study of Language and Information Publications.

Institute of Research in Cognitive Science
3401 Walnut St., Suite 400A
Philadelphia, PA 19104-6228
elenimi@babel.ling.upenn.edu