Joint Information Value of Syntactic and Semantic Prominence for Subsequent Pronominal Reference

Ralph L. Rose

Faculty of International Communication
Gunma Prefectural Women's University
Gunma, Japan
`rose@gpwu.ac.jp`

Abstract

Entities realized in prominent syntactic positions receive some preferential treatment when referred to in a subsequent discourse segment: In particular, they are preferentially referred to with reduced referring expressions, often pronouns. This has been observed in both production (Arnold, 1998) and perception (Gordon et al., 1993; Almor, 1999). As a result, syntactic prominence has been regarded as a primary factor in determining the salience of entities. However, in English, syntactic role and semantic role are often conflated. That is, entities realized as syntactic subjects are often semantic agents while those realized as objects are often patients. Thus, it could be that salience is determined by semantic prominence rather than syntactic prominence. In this paper, I present some evidence from a corpus investigation comparing the relative contribution of syntactic and semantic prominence to the salience of entities for subsequent pronominal reference. Using the concept of the value of information from Information Theory (Shannon, 1948), results suggest that syntactic and semantic prominence are comparably informative, but that together they are more informative than either is alone.

1 Introduction

Many studies of discourse production and perception have observed that entities evoked in subject position are treated somewhat differently than those evoked in other positions when those entities are referred to subsequently. For instance, consider the short discourse in (1).

(1)   a. $Luke_i$ hit Max_j.
       b. Then $he_{i/\#j}$ ran home.
       b'. Then #Luke/Max ran home.

While the pronoun in (1b) is ambiguous and could be interpreted as referring to either LUKE or MAX, the preferred interpretation is LUKE, the subject of the preceding sentence (cf., Hudson-D'Zmura and Tanenhaus (1997); Mathews and Chodorow (1988)). Similarly, repeated reference to LUKE by name as in (1b') is more marked than repeated reference to MAX by name (Gordon et al., 1993; Almor, 1999). These observations are from the hearer's perspective, but even from the speaker's perspective, similar preferences have been observed. Brown (1983) observed that entities introduced as subjects persisted longer than those introduced in other syntactic positions: That is, there were more contiguous utterances in which the entity was referred to again.

Many models of discourse production and processing capture these observations through two assumptions. First, the *salience* of entities evoked in a discourse determines how subsequent reference to those entities should be performed (Ariel, 1988; Gundel et al., 1993). Second, syntactic information is a primary or even sole factor which deter-

mines salience (Grosz et al., 1995; Lappin and Leass, 1994). Thus, according to this kind of model, the first sentence in (1) introduces two entities into the discourse representation, LUKE and MAX. With respect to the syntactic prominence hierarchy shown in (2), in this representation, LUKE is more salient because it was realized in subject position while MAX is less salient having been realized in object position.

(2) $subject > object > oblique > none$

One problem with this account is that in such languages as English, syntactic information is often conflated with semantic information. That is, syntactic subjects are often semantic agents and carry more PROTO-agent entailments (e.g., sentience, volition; Dowty (1991)), while syntactic objects are often semantic patients and carry more PROTO-patient entailments (e.g., undergo change-of-state, causally affected). Thus, assuming a semantic prominence hierarchy as in (3) (cf., thematic hierarchies in Jackendoff (1972), Speas (1990)), it could be the case that LUKE is more salient than MAX in (1a) not because it is realized in subject position, but rather because it is realized as an agent (of the hitting event).

(3) $agent > patient > others$

The purpose of this paper is to investigate this possibility. To restate it as a question: What is the relative contribution of syntactic prominence and semantic prominence to the salience of entities evoked in a discourse? I investigate this question with a corpus investigation which looks at coreference across adjacent utterances and the form of referring expression (pronoun or description) used in subsequent reference. The results are presented in information-theoretic terms (Shannon, 1948) and suggest that while syntactic and semantic prominence are comparably informative about the form of subsequent reference, taken together, syntactic and semantic prominence are more informative than either is alone.

In the next section, I describe the basic discourse model I assume in this paper and then in Section 3 I describe the corpus used in this study. Section 4 contains an overview of information theory and particularly the concept of the value of information. I report the results of the study in Section 5 along with interleaved discussion.

2   Discourse Model

In this paper, I assume a model of discourse processing in which the current utterance is processed with respect to the context; that is, the representation of the discourse so far (Kamp and Reyle, 1993; Kehler, 2002). I assume that the context contains representations of the entities evoked in the discourse. Following Karttunen (1976) and Heim (1982; 1983), I call them *discourse referents*. The set of referents is a partially-ordered list, the order of which is determined by a number of factors including syntactic role and recency (see Hirst (1981) and Mitkov (2002) for an overview of these and many other factors). I take the highest ranking referent to be the most salient referent in the current context. As such, if this referent is evoked in the current utterance, then it should be done so pronominally (cf., Rule 2 of Centering Theory, Grosz, et al., (1995)). This then is a useful metric for determining which referents in the context are more salient than others.

This is the approach I use in the corpus analysis in order to examine which referents are most salient and subsequently which syntactic and semantic features are most informative for determining their salience. However one simplification I make is to assume that recency determines that all referents evoked in the most recent utterance are more salient than those evoked in earlier utterances. Thus, while inter-utterance coreference could conceivably span multiple utterances, the present study only considers coreference in adjacent utterances.

The theoretical approach which I take in this study embodies the speaker's point of view in discourse processing. In other words, I am investigating what the speaker takes as salient in the discourse and the encoding decisions made as a result of that. However, I take salience to be a feature of discourse representation which is ultimately used by both hearer and speaker in their respective tasks. The precise way in which each uses salience may be different, but I assume that they rely on the same core notion of salience in the process of discourse production or perception (cf., Prince (1986) and Blutner (1998; 2000)).

3 Corpus Design

The corpus is composed of texts selected from *InterText* (`http://www.intertext.com`)—an online, refereed magazine of fiction. At present the corpus contains five complete texts of varying lengths comprising a total of 5,480 words. The selected texts are third-person narratives with minimal quoted passages. These texts were manually marked-up using XML. In this section, I describe the relevant markup elements and how the corpus was analyzed in order to answer the main research question.[1]

3.1 Utterances

The texts were first parsed into sentence nodes, <s>, based on their appearance in the text: word strings terminated by a period (except of course for periods marking an abbreviation). The <s> nodes were further marked with a relatively shallow parse based on clause relations. Each clause, <c>, contained at most one <verb> child. The noun-phrase, <np>, and clausal arguments of a verb were marked as siblings of the <verb>. The text shown in (4) was thus tagged as in (5) (leaving out currently irrelevant details).

(4)   John hit Matt. He told his teacher that John did so.

(5)
```
<s>
 <c>John hit Matt</c>
 <punc>.</punc>
</s>
<s>
 <c>He told his teacher that
  <c>John did so</c>
 </c>
 <punc>.</punc>
</s>
```

In the analyses which follow, I will be investigating instances of inter-utterance coreference. In terms of the corpus, I define an utterance as a <c> node which is the immediate child of a <s> node. Thus, the embedded clause in (4), *John did so* is not an utterance. On the other hand, conjoined clauses (e.g., *The building is tall and it is old.*) are treated as separate utterances.

---

[1] It is important to note here that the corpus mark-up has been performed entirely by myself. Thus, at present there is no inter-rater validation. However, numerous passes over the corpus by me have likely ensured a high degree of intra-rater consistency.

One final note here is that this study looks only at coreference between noun-phrases. Thus such things as event references as in *John secretly pinched Matt but the teacher saw it* are not included. It is doubtful that this exclusion has much effect on the overall results since there were only a handful of such cases in the corpus.

3.2 Syntactic Information

The syntactic role of each argument <np> was marked as "subject", "object", or "oblique". Any other <np> nodes which were not arguments of a verb were marked as "none" (i.e., not subject, object, or oblique). In each clause, the nearest <np> node preceding the <verb> was marked as the subject; the nearest <np> node following the <verb> but not immediately preceded by a preposition was marked as the object (so-called double-object constructions like *give Mark the pen* were marked with two objects); and any (remaining) <np> node immediately preceded by a preposition was marked as an oblique. Thus, (6) was tagged as in (7).

(6)   Ken threw the frisbee to Jaime.

(7)
```
<s>
 <c>
  <np synrole="subject">
   Ken</np>
  <verb>threw</verb>
  <np synrole="object">
   the frisbee</np>
  to
  <np synrole="oblique">
   Jaime</np>
 </c>
 <punc>.</punc>
</s>
```

3.3 Semantic Information

The semantic role of each <np> argument was marked with respect to two semantic systems: the FrameNet (Baker et al., 1998) system of frames and elements and the PROTO-role entailments of Dowty (1991). Here, I briefly explain each of these.

3.3.1 FrameNet

Based on the frame semantics of Fillmore (1968; 1976), the FrameNet system defines a large number of conceptual frames (e.g., intentionally_affect, transitive_action), each of which incorporates a set of frame elements (i.e., thematic roles: agent, patient, etc.) which participate in that frame. Each

frame encompasses a number of lexical items which invoke that frame and therefore define the particular roles that the arguments of each item play. For instance, the verb *throw* invokes the cause_motion frame and therefore takes several participants including an agent, a theme, and a goal.

In the present study, the semantic role of each <np> argument of a <verb> was determined by consulting the FrameNet database for the frame which encompassed that verb and then assigning the respective frame element labels to the <np> nodes. If a verb was not in the FrameNet database, then the database was searched for a suitable alternative (e.g., via synonymic or hypernymic relations). Thus, the sentence in (6) was tagged as in (8).

(8)
```
<s>
 <c>
  <np semrole="agent">
    Ken</np>
  <verb>threw</verb>
  <np semrole="theme">
    the frisbee</np>
  to
  <np semrole="goal">
    Jaime</np>
 </c>
 <punc>.</punc>
</s>
```

3.3.2 PROTO-roles

Dowty (1991) proposes an alternative view of the linking between lexical conceptual structure and syntax through semantic entailments placed on arguments by a verb. He posits two sets of PROTO-role entailments as in (9).

(9)   PROTO-agent entailments

- sentience
- volition
- cause event or change-of-state
- undergo movement

PROTO-patient entailments

- undergo change-of-state
- causally affected
- incremental theme
- stationary

Under Dowty's theory, arguments of a verb may carry any number of these entailments. A selection principle then determines that the argument which carries the most PROTO-agent entailments becomes the surface subject. The remaining argument with the most PROTO-patient entailments becomes the object. Any other arguments become obliques. It is important to notice then that under this system, arguments may take on the PROTO-agent or PROTO-patient roles in varying degrees. With one verb, the argument realized as subject may carry all four PROTO-agent entailments while with another verb, the argument realized as subject may carry only one or two. Furthermore, some crossover between the roles is possible: An argument realized as a subject may carry some PROTO-patient entailments while an argument realized as an object may carry some PROTO-agent entailments.

In the corpus, PROTO-agent entailments for every <np> argument were marked. The entailments associated with any particular verb were determined using a series of linguistic tests described in Rose (2005). Thus, (6) was marked as in (10).

(10)
```
<s>
 <c>
  <np sentience="yes"
      volition="yes"
      stationary="yes">
    Ken</np>
  <verb>threw</verb>
  <np movement="yes">
    the frisbee</np>
  to
  <np stationary="yes">
    Jaime</np>
 </c>
 <punc>.</punc>
</s>
```

3.3.3 FrameNet vs. PROTO-roles

The two different semantic systems used in this study provide an interesting contrast. In Frame Semantics, upon which FrameNet is based, case roles are seen as derived from primitive, psychologically real semantic concepts (Fillmore, 1968). PROTO-roles, on the other hand, are seen merely as labels for flexible configurations of semantic entailments (Dowty, 1991). If one or the other of these two views could be shown as more closely linked to salience, this may suggest different things about the nature of salience. For instance, if the FrameNet approach can be shown to be better, this may suggest an interesting link between salience and semantic primitives via the roles that entities are seen to play in conceptual frames.

3.4 Coreference Information

In order to be able to examine coreference relationships across adjacent utterances, every referential noun phrase (i.e., excluding such things as expletive *it*) was marked with an identifier string. Within any given text, all noun-phrases which were interpreted as referring to the same real-word referent were given the same identifier. Thus, (11) was marked as shown in (12).

(11) Louis watched a ballerina. She was graceful.

(12)
```
<s>
  <c>
   <np id="LOUIS">Louis</np>
   <verb>watched</verb>
   <np id="BALLERINA">
      a ballerina</np>
  </c>
  <punc>.</punc>
<s>
<s>
  <c>
   <np id="BALLERINA">She</np>
   <verb>was</verb>
   graceful
  </c>
  <punc>.</punc>
</s>
```

4 Information Theory

The corpus analysis which follows makes use of one fundamental concept in Information Theory (Shannon, 1948): the value of information (hereafter, EIV). EIV is based on the entropy, H—an estimate of the uncertainty of the outcome—of a given probability space. H for a probability space with N possible outcomes can be calculated as shown in (13) where $P(n)$ is the probability of the n-th outcome.

(13) $-\sum_{n=1}^{N} P(n)log_2 P(n)$

For a given question in which all possible outcomes are equally likely (e.g., the flip of a fair coin), the entropy is very high. However, if we learn some information, x, that causes one outcome to be far more likely to occur, then our uncertainty will decrease: H will be reduced. The amount of entropy reduction as a result of learning x, $H_r(x)$, is thus calculated as the difference between the initial entropy, H, and the conditional entropy $H(x)$ (i.e., H given x).[2]

---

[2]The value amounting to the reduction in entropy has also been referred as the *entropy value* (van Rooy, 2004).

To illustrate, consider the following problem: If I open a novel to a random page and point to a random letter on the page, what is the probability, P, that the letter is 'u'? Without any other information, P is simply the prior probability of the occurrence of 'u' in the language as a whole. Using this prior probability we could calculate the entropy, H, of the problem. However, imagine we learn that the preceding letter is 'q'. Then we can be much more certain that the letter in question is 'u'. Thus, the conditional entropy, $H('q')$, will be less—a reduction in entropy.

Entropy reduction may be either positive or negative: learning that x is true may make us more certain while learning that x is false may make us less certain about some outcome. It is therefore useful to calculate the value of learning whether or not x is true. In other words, it is useful to know what is the overall value of asking the question of whether x is true or false. In information theory, this value is estimated as the weighted sum of the entropy reductions for all possible outcomes of x (here, true or false). This value is known as the estimated information value, EIV. Formally, the EIV of learning whether or not x is calculated using the formula shown in (14), where $P(x)$ is the prior probability of the occurrence of x.

(14) $EIV(x) = P(x)H_r(x) + P(\neg x)H_r(\neg x)$

A good illustration of information value comes from the game "Who am I?" in which one person pretends to be some famous person and others must ask yes/no questions to find the identity. In this scenario, what is an informative (i.e., having a large information value) first question assuming that there is no bias in the choice of famous person? One candidate would be "Are you a male/female?" In this case, both terms in the sum of (14) will be at a maximum and thus EIV will be large. However, a question like "Are you Albert Einstein?" will be much less informative: While the entropy reduction if the answer is yes, $H_r(x)$, is large, the probability the answer is yes, $P(x)$, is very small. If the answer is no then the converse is true. Thus both terms in the sum of (14) will be small and EIV will be small. Of course, if after several questions we have learned that the mystery person is male, is a scientist, lived

in the 20th century, and won a Nobel Prize, then the EIV would be much larger.

In the present study, I am investigating the information value of syntactic and semantic prominence toward determining the salience of discourse referents. This is done by asking, for example, the following question: What is the information value of learning whether or not a particular discourse referent was a subject to the probability of its being pronominalized in subsequent reference? This information value, $EIV(subject)$, can be calculated using the formulas above. Likewise, the information values for the other syntactic and semantic features can be calculated. Finally, I will calculate the net information value, EIV_{tot}, for syntactic prominence as the total of the EIVs for the various syntactic features (i.e., $EIV(subject)$, $EIV(object)$, etc.). Similarly, I will calculate the EIV_{tot} for semantic prominence as the total of the EIVs for the various semantic features. Therefore, the central question becomes whether either information about syntactic prominence or semantic prominence is more informative (i.e., larger EIV_{tot}) than the other or if they are equally informative. A second question is whether syntactic and semantic information *together* is more informative than either is alone. These two questions are formally summarized in (15)-(16).

(15) Is the syntactic prominence EIV_{tot} greater than, equal to, or less than the semantic prominence EIV_{tot}?

(16) Is the joint syntactic and semantic prominence EIV_{tot} greater than either the syntactic or semantic prominence EIV_{tot}?

With respect to (15), if results show that syntactic and semantic prominence are equally informative, then another question may be posed: Are syntactic and semantic prominence redundant with each other or are they at least somewhat independent but equally informative? An answer to this question may be found by looking at the answer to (16). If the joint information value is higher than either is alone, then they cannot be redundant and must therefore be independent.

5 Results and Discussion

In the corpus there are 291 cases of inter-utterance coreference. In 224 (77%) of these coreference cases, the coreferent noun phrase in the latter utterance is pronominalized. Thus, the entropy of pronominalization can is calculated as shown in (17) where $P(pro)$ is the probability of pronominalization.

$$
\begin{aligned}
(17)\quad H &= -[P(pro) * log_2 P(pro) + \\
&\quad\ P(\neg pro) * log_2 P(\neg pro)] \\
H &= -[224/291 * log_2(224/291) + \\
&\quad\ 67/291 * log_2(67/291)] \\
H &= 0.778
\end{aligned}
$$

This value serves as the baseline for entropy reduction: How much is entropy reduced from $H = 0.778$ by learning some information about syntactic or semantic prominence? In this section, I will present these results along with some interleaved discussion. However, before presenting the results, it is necessary to deal with one complication. The referents in the current context may have been realized in multiple syntactic positions and semantic roles. For instance, in (18), as a verbal argument, JOHN has been realized as a subject and an object, an experiencer and a recipient, and carries the entailments sentience, volition, and stationary.

```
(18)  <s>
       <c>
        <np id="JOHN"
            synrole="subject"
            semrole="experiencer"
            sentience="yes"
            volition="yes">
          John</np>
        <verb>wants</verb>
        <c>
         <np>his father</np>
         to
         <verb>give</verb>
         <np id="JOHN"
             synrole="object"
             semrole="recipient"
             stationary="yes">
           him</np>
         <np>a bicycle</np>
        </c>
       </c>
       <punc>.</punc>
      </s>
```

In short, there is an overlap of information caused by such co-occurrences. It seems likely that these co-occurrences are not independent of one another, but accounting for these dependencies requires a rather sophisticated mathematical model.

For the present research, I will therefore make certain simplifying assumptions about syntactic prominence and the two semantic prominence approaches. These assumptions will be clarified in greater detail in the respective sections below.

5.1 Syntactic Prominence

For syntactic prominence information, I assume that for any given referent, the role highest on the syntactic hierarchy shown in (2) determines that referent's salience. Given this, the results shown in Table 1 indicate that learning that a referent was realized as a subject is much more informative than learning it was realized in any other role about whether or not subsequent reference to that referent will be pronominalized or not.

Table 1: Information Value of Syntactic Prominence

x	$EIV(x)$
subject	0.059
object	0.021
oblique	0.010
none	0.011
EIV_{tot}	0.101

The idea that the information value of subjecthood is much higher than that of other syntactic roles is especially interesting in that it resembles the binary nature of many information-packaging theories (e.g., *topic-comment* in (Gundel, 1974) and *topic-focus* in (Sgall, 1967)). The concept of the value of information may may provide a useful of quantifying these theories.

5.2 Semantic Prominence

5.2.1 FrameNet Roles

In the corpus, 158 different frame elements occur. Here, I collapse these elements into seven groups as shown in (19). Each group is shown with a word that briefly describes the central property of the elements in that group as well as some examples of elements in that group.

(19) **1 - agentivity** agent, deformer, driver
      2 - perception cognizer, experiencer
      3 - movement theme, impactor, message
      4 - affected created_entity, victim
      5 - movement parameters direction, ground
      6 - events activity, event
      7 - other specifier, none

The ordering of the groups shown parallels orderings given in thematic hierarchies proposed in the literature on syntactic linking theories (cf., Jackendoff (1972; 1990) and Speas (1990)). Similar to the simplifying technique for syntactic information above, for a given referent, the one of its semantic roles which is highest on this hierarchy is regarded as the role which determines the salience of that referent. The results are thus shown in Table 2.

Table 2: Information Value of Semantic Prominence via FrameNet Roles

group	$EIV(group)$
1	0.013
2	0.045
3	0.012
4	0.002
5	0.005
6	0.004
7	0.019
EIV_{tot}	0.101

Two results are notable. First, it is interesting that the perception roles in group 2 are more informative than the agentive roles in group 1, in spite of the fact that agentive roles are usually posited to be highest on many thematic hierarchies. This suggests that sentience is more important to the salience of entities evoked in a discourse than agentivity. This would seem to parallel other results showing the importance of animacy to the salience of discourse entities (Prat-Sala and Branigan, 1999).

The second interesting result is that the total information value of semantic prominence under FrameNet is equal to that of syntactic prominence. I will discuss the implications of this below. Before that, I present the results for the other semantic prominence approach used in this study.

5.2.2 PROTO-roles

A particular discourse referent may carry more than one PROTO-role entailment. In order to avoid the overlap problems that this generates in the

present analysis, I use a simple transformation. For each referent, I calculate a parameter I call PROTO-agency as the total number of (unique) PROTO-agent entailments on that referent minus the total number of PROTO-patient entailments. Thus, PROTO-agency ranges in integer values from +4 to -4 (although in this corpus, there were no instances of -4). Under this transformation, the results are as shown in (3).

Table 3: Information Value of Semantic Prominence via PROTO-role entailments

PROTO-agency	EIV(PROTO-agency)
+4	0.000
+3	0.001
+2	0.045
+1	0.005
0	0.045
-1	0.000
-2	0.001
-3	0.002
-4	***
EIV_{tot}	0.098

Results here show that semantic prominence with respect to PROTO-roles is comparably informative to the semantic prominence with respect to FrameNet as well as to syntactic prominence.

5.3 Joint Information Value

While the above results have looked at the information value of learning about the syntactic or semantic prominence of a referent (i.e., learning that it was realized as a subject or as a group 1 FrameNet element or with a PROTO-agency of +4 or so on), in this section, I look at the value of learning some joint information. That is, what is the value of learning that a referent was realized as, say, a subject *and* a group 1 role?

The fact that the EIV_{tot} values of syntactic and semantic prominence are essentially equal suggests that either they are essentially redundant or that they are at least somewhat independent, but comparably informative. If the former is the case, then the joint information value should be no different than that of each alone. However, if there is some independence between the two pieces of information, then the joint information value may increase.

The joint information value of syntactic and semantic prominence was calculated by crossing the four syntactic roles against the seven FrameNet groups or the nine levels of PROTO-agency, and then calculating the EIV for each of the pairings (e.g., subject/group 1, subject/group 2, etc.). The total information value, EIV_{tot} was then calculated as the total of these individual $EIVs$. The final results are shown in Table 4.

Table 4: Joint Information Value of Syntactic and Semantic Prominence

	EIV_{tot}
syntactic role $\times$ FrameNet group	0.165
syntactic role $\times$ PROTO-agency	0.141

The joint information value of syntactic and semantic prominence is higher than that of either factor alone. Thus, the results suggest that syntactic and semantic prominence are not redundant with each other and that each provides at least some unique information with respect to the pronominalization of subsequent reference. This conclusion should be regarded as tentative, however, because the differences noted above have not been statistically confirmed.[3]

One interesting result here, though, is the fact that the PROTO-role information is not quite as informative as the FrameNet group information when taken together with syntactic role. This, however, could be a by-product of the transformation on the PROTO-role information described in Section 5.2.2.

---

[3]I have been unable to find established procedures for the statistical evaluation of EIV values. I have thus attempted two different procedures. In the first, the variances of the mean were estimated through a series of calculations paralleling those of the EIV calculations. Under these procedures, the differences between the joint EIVs and individual EIVs are significantly different at the $\alpha = 0.05$ level. However, this conclusion is suspect because the procedure assumes normal distributions throughout. It is doubtful that this is the case since, for example, even if x is normally distributed, $log_2(x)$ is not.

The second procedure is a bootstrap in which the original sample of inter-utterance coreference instances was resampled with replacement 10,000 times. Under this procedure, the differences between the joint EIVs and the individual EIVs was not shown to be significant. However, even this procedure is suspect because measures of skewness and kurtosis show that the bootstrap distribution is nonnormal.

I am currently in search of more reliable and valid statistical procedures in order to evaluate the results. For the present paper, I discuss the results as if the differences are real, but refrain from making statistical claims about the differences.

This transformation is a mathematical convenience and glosses over semantic distinctions between the various entailments. Perhaps a more sophisticated transformation would result in a greater information value.

6 Conclusion

Under the discourse model presented above, the results presented in this corpus analysis suggest that the salience of referents in a discourse is influenced by both syntactic and semantic information: Taking both into account results in greater predictive ability for the form of subsequent reference. These results are thus in line with a view of discourse processing in which salience represents information about discourse structure: the more salient a referent is in the current context, the greater the information value about the structure of subsequent discourse, particularly the form of referring expressions. Information theory thus potentially offers another view of the relative value of the different factors known to affect discourse salience and may provide another means by which to narrow down on which factors are most crucial.

The fact that syntactic and semantic information seem to be at least partly independent in their influence on salience suggest that models of discourse salience should include some account of semantic information as distinct from syntactic information. This is especially relevant to modular approaches in which one module is responsible for structure while an independent module is responsible for interpretation. The results here may be relevant for determining how these modules interact for the purpose of determining salience.

The improvement in the joint information value suggests that computational implementations of discourse salience models might see some improvement by the inclusion of semantic prominence information. For instance, if the assumption that salience is a core notion common to both speaker and hearer is correct, then the present results would indicate that pronoun resolution algorithms might also benefit from the inclusion of semantic prominence as a contributing factor.

In this study, two different semantic systems were employed to evaluate semantic prominence. The joint information values suggest that the FrameNet system may be more informative than the PROTO-role system. However, as noted above, this difference may not be real. If it is real, then an interesting line of future investigation would be to look more closely at the relationship between salience and the notion of primitive semantic roles as assumed in frame semantics. On the other hand, if the difference between the two systems turns out not to be real, then there is a practical conclusion to make: Technologically speaking, the PROTO-role system is less cumbersome than the vast network of frames and roles in FrameNet and therefore may be more efficient in the implementation of mechanisms for discourse processing and salience.

Further investigation of the role of syntactic and semantic prominence along the lines presented here might include looking at different languages. In English—the language used in this study—syntactic and semantic role is often conflated as noted early in this paper. However, in languages where word order is more free such as Spanish or Japanese, the distinction between syntactic and semantic prominence may be easier to observe. Such work may provide a clearer view of the degree to which syntactic and semantic prominence each determine the salience of discourse referents.

References

Amit Almor. 1999. Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological Review*, 106:748–765.

Mira Ariel. 1988. Referring and accessibility. *Journal of Linguistics*, 24:65–87.

Jennifer Arnold. 1998. Marking salience: The similarity of topic and focus. Unpublished Manuscript.

Collin Baker, Charles Fillmore, and John Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the COLING-ACL*.

Reinhard Blutner. 1998. Lexical pragmatics. *Journal of Semantics*, 15:115–162.

Reinhard Blutner. 2000. Some aspects of optimality in natural language interpretation. *Journal of Semantics*, 17:189–216.

Cheryl Brown. 1983. Topic continuity in written English narrative. In Talmy Givón, editor, *Topic Continuity*

in Discourse: A Quantitative Cross-Language Study, pages 313–342. John Benjamins, Amsterdam.

David Dowty. 1991. Thematic proto-roles and argument selection. *Language*, 67:547–619.

Charles Fillmore. 1968. The case for case. In Emmon Bach and Robert Harms, editors, *Universals in Linguistic Theory*, pages 1–90. Holt, Rhinehart and Winston, New York.

Charles Fillmore. 1976. Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech*, volume 280, pages 20–32.

Peter Gordon, Barbara Grosz, and Laura Gilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17:311–347.

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

Jeanette Gundel, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions. *Language*, 69:274–307.

Jeanette Gundel. 1974. *The Role of Topic and Comment in Linguistic Theory*. Ph.D. thesis, University of Texas, Austin.

Irene Heim. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts, Amherst.

Irene Heim. 1983. File change semantics and the familiarity theory of definiteness. In R. Bauerle, Ch. Schwarze, and A. Von Stechow, editors, *Meaning, Use, and Interpretation*, pages 164–189. DeGruyter, Berlin.

Graeme Hirst. 1981. *Anaphora in Natural Language Understanding: A Survey*. Springer-Verlag, Berlin.

Susan Hudson-D'Zmura and Michael Tanenhaus. 1997. Assigning antecedents to ambiguous pronouns: The role of the center of attention as the default assignment. In M. Walker, A. Joshi, and Ellen Prince, editors, *Centering Theory in Discourse*, pages 199–226. Clarendon Press, Oxford.

Ray Jackendoff. 1972. *Semantic Interpretation in Generative Grammar*. MIT Press, Cambridge, MA.

Ray Jackendoff. 1990. *Semantic Structures*. MIT Press, Cambridge, Massachusetts.

Hans Kamp and Uwe Reyle. 1993. *From Discourse to Logic*. Kluwer Academic, Dordrecht.

Lauri Karttunen. 1976. Discourse referents. In James McCawley, editor, *Syntax and Semantics, Vol. 7: Notes from the Linguistic Underground*, pages 363–385. Academic Press, New York.

Andrew Kehler. 2002. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, Stanford University, CA.

Shalom Lappin and Herbert Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20:535–561.

Alison Mathews and Martin Chodorow. 1988. Pronoun resolution in two-clause sentences: Effects of ambiguity, antecedent location, and depth of embedding. *Journal of Memory and Language*, 27:245–260.

Ruslan Mitkov. 2002. *Anaphora Resolution*. Longman, London.

Mercè Prat-Sala and Holly Branigan. 1999. Discourse constraints on syntactic processing in langauge production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, 42:168–182.

Ellen Prince. 1986. On the syntactic marking of presupposed open propositions. In *Papers from the Parasession on Pragmatics and Grammatical Theory, 22nd Regional Meeting, Chicago Linguistic Society*, pages 208–222, Chicago, IL.

Ralph Rose. 2005. *The Relative Contribution of Syntactic and Semantic Prominence to the Salience of Discourse Entities*. Ph.D. thesis, Northwestern University.

Petr Sgall. 1967. Functional sentence perspective in a generative description. *Prague Studies in Mathematical Linguistics*, 2:203–225.

Claude Shannon. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656.

Margaret Speas. 1990. *Phrase Structure in Natural Language*. Kluwer Academic Publishers, Dordrecht, Netherlands.

Robert van Rooy. 2004. Relevance and bidirectional optimality theory. In Reinhard Blutner and Henk Zeevat, editors, *Optimality Theory and Pragmatics*, pages 173–210. Palgrave Macmillan, Oxford.