

Identifying Attentional Center in the Light of Saliency Weight: A Key Issue in Anaphora Resolution

Dafang Wang

School of Foreign Languages, Renmin University of China

No.59 Zhongguancun Street, Beijing, China

Tel: 86-134-3932-2999 E-mail: wangdafang@ruc.edu.cn

Received: July 25, 2011

Accepted: August 12, 2011

Published: October 1, 2011

doi:10.5430/wjel.v1n2p9

URL: <http://dx.doi.org/10.5430/wjel.v1n2p9>

Abstract

Existing accounts on the relationship between discourse saliency and choice of referring expression differ in what they view as relevant factors that establish entities as more or less salient. In this paper, saliency is treated as a gradient and relative cognitive property reflected by the partial order among a set of candidate referents in a dynamically updated attentional state. To be specific, I will focus on three linguistic properties that have been claimed to influence the saliency weight in English: *information status*, *referential form*, and *empathy*. Along with the systematic investigation, I will clarify the working definition of some concepts and briefly introduce the linguistic diagnostics adopted for saliency assessing and ranking. It is hoped that the efforts made here will shed light on the large-scale empirical work that remains to be done in the near future.

Keywords: Saliency weight, Attentional center, Anaphora resolution

1. Introduction

There exists a general consensus that the most salient referents--those referents that are currently at the centre of attention and most prominent at that point in the discourse -- are referred to with the most reduced referring expressions, in most cases a (zero) pronoun, depending on the language. Such a correlation has been encoded in many prior works on anaphora resolution: Givón (1983) associates a gradient scale of *topicality* with choice of linguistic expression, with the most topical entity predicted to be referenced with a zero pronoun when permitted by the language; Ariel's (1990) *Accessibility Marked Scale* associates null forms and pronouns with entities that fall on the high end of the accessibility scale; the *Givenness Hierarchy* proposed by Gundel et al. (1993) associates entities at the in focus end of the hierarchy with pronominal reference.

Here comes the crucial question: what kinds of factors influence a referent's saliency and make it a good candidate to be referred to with a reduced anaphoric expression? Treating saliency as a gradient and relative cognitive property reflected by the partial order among a set of candidate referents in a dynamically updated attentional state, this paper will focus on three linguistic properties that have been claimed to influence the saliency weight: 1) *information status*, 2) *referential form*, and 3) *empathy*. Along with the systematic investigation, I will clarify the working definition of some concepts and briefly introduce the linguistic diagnostics adopted in this study for saliency assessing and ranking.

2. Dynamic Loading of Information: Topic vs. Focus

A cognitively prominent antecedent is more strongly activated than a non-prominent antecedent within a mental model of the discourse. This suggests that knowledge of information structure is immediately available upon encountering a pronoun. To clarify potential sources of cognitive prominence for different discourse entities, theories of information structure may provide us with enlightenment and guidance. Such theories enable us to analyze the informational role that different parts of an utterance play, determining which parts of the sentence present new or unpredictable information and which present referents to which such new information can be added. There are two central information-structure categories: **Topic** and **Focus**. Refraining from a review of the vast literature on these two categories, I will instead treat them as two complementary notions in this paper. In other words, one of them can be defined as what the other is not. Generally speaking, the topic provides the context for the main predication, and the focus presents information that is unpredictable and applicable to the topic. Although each has independent properties, both Topic and Focus are often characterized as the salient part of an utterance. To explain how, we will discuss these two categories in more detail.

2.1 Topic

It is generally agreed upon that **Topic** can be defined in terms of the “aboutness” condition (Reinhart, 1982). To facilitate future analysis, I will make a distinction here between *Local Topic* (LT) and *Global Topic* (GT). At the local level, topic is the (presupposed) information that the rest of the assertion in an utterance modifies; and more particularly, it often refers to a linguistic constituent of the category NP. In this sense, LT is defined in terms of the referential givenness status of entities. It is usually realized by a referring expression and designates a referent identified within a discourse (Lambrecht, 1994). From the angle of functional syntax, LT can be roughly equated with the topical theme in Halliday’s grammatical system.

In comparison, GT is a more complex notion as it functions at a higher level. Defined in an analogous fashion to that of LT, GT is what a set of propositions is about at a given point in discourse. From the perspective of anaphora processing, GT here is assumed as a referent that has been the local topic for more than one sentence without crucial intervening local topics within a discourse (segment). In this way, the identification of GT becomes much more manageable because it is based upon the topical analysis at the sentence level. As a matter of fact, the continuity of LT is an important criterion in discourse segmentation with GT serving as a salient filing label.

It should be noted that these two concepts often coincide within a discourse segment, but not necessarily so. For example, in (1), the GT is *Mr. Bennet*. It coincides with the LT of (1a), but differs from that of (1b), where the LT shifts to Mr Bennet’s *affection* for his daughter Elizabeth. This kind of part-whole relation is fairly common in the topical construction of discourse. As the current study focuses on coreferential pronouns, we will not probe into this kind of topic shift in detail.

(1) a. **Mr. Bennet** missed his second daughter exceedingly.

b. **His affection for her** drew him oftener from home than any thing else could do.

(Jane Austen: *Pride and Prejudice*)

Regarding the influence of topic status on referential choice, many scholars (Crawley et al., 1990; Marslen-Wilson et al., 1993; Stevenson et al., 1995; Arnold et al., 2000 etc.) use anaphoric elements as a window to gain insight into what makes referents salient, and thus suggest or justify a preference for co-reference of a pronoun with a topical entity. Here comes the practical question of topic identification. It might be argued that English is impoverished, compared to other languages, with regard to the marking of topics. It does not have a special morphological marker, such as *-wa* in Japanese and *-i(n)* in Korean (cf. Li & Thompson for a survey of languages with and without explicit topic marking). With respect to anaphora resolution, I select several identification methods of Local Topic for practical reasons. They are listed below with incremental markedness:

2.1.1 Subjecthood

In English, subject has special properties not shared by other sentence constituents, such as determination of the person and number features exhibited by the verb of a finite clause; preverbal position in finite clauses; the privilege of being omitted in the infinitive, in gerund clauses, and in conjoined finite clauses, and so on. Thus in the absence of stronger salience-indicators of other discourse entities, **grammatical subject** will be perceived as the DEFAULT topic. As mentioned above, this tenet has been widely employed in automatic algorithms for anaphora processing. What should be noted is that **expletive subjects** like *there* in example (2) and *it* in (3) are not regarded as topics. This is because topics need to be referring expressions in the current study, and must designate a referent identified within a discourse. In this kind of cases, the complementing constituent will be treated as the topic instead, like *my aunt* in (2) and *Squealer* in (3), which turn out to be the antecedent of the pronouns *she* and *his/he* in the upcoming discourse.

(2) And **there** was my aunt, all the time I was dressing, preaching and talking away just as if she was reading a sermon.
(Jane Austen: *Pride and Prejudice*)

(3) Yes, **it** was Squealer. A litter awkwardly, as though not quite used to supporting his considerable bulk in that position, but with perfect balance, he was strolling across the yard. (George Orwell: *Animal Farm*)

2.1.2 Topic-denoting syntactic constructions

Since the early Prague School work on syntax and discourse function (e.g. Firbas, 1966, 1992), researchers have accumulated evidence for a correlation between sentence position and information status in the discourse. In particular, it has been noticed that preposing constructions which place canonically postverbal constituents in preverbal position always mark the preposed information as topic within the discourse. As shown in (4), it is a common observation that corresponding active and passive sentences differ in what they appear to be ‘about’ even though they describe the same state of affairs. Besides **Passive Construction** which can be covered by the subjecthood condition, the syntactic

structures most widely and consistently associated with topic denoting are **Topicalization** and **Left-dislocation**, both involving a topical NP adjunct to the left of a clause, and some kind of referential link between the NP adjunct and a constituent of the clause. In Topicalization, the link between the NP adjunct and the full clause is an obligatorily co-indexed gap, as in (5); while in Left-dislocation, the coreferential link is embodied by a pronoun or other anaphorically interpreted element, as in (6). Another effective topic marking construction is **Right-dislocation**, which is akin to Left-dislocation, simply dislocates the NP adjunct to the right of a clause, like (7). The major discourse function of Right-dislocation constructions is to enable the addressor to signal to the addressee a shift in attention and refer to entities that are salient in the global discourse situation rather than in the current forward-looking center. A still more marked syntactic construction which yields an interesting test for syntactic markedness and topic properties of particular NP is **Raising-structure**, which has two subtypes: **Raising to Object Structure (ROS)** and **Raising to Subject Structure (RSS)**. In ROS, as illustrated in (8), the grammatical role of *poinsettias* varies from the lower-clause subject in (8a) to the clausal object in (8b). In RSS, as illustrated in (9), *light energy* is raised to be the subject in (9b) though it is actually the subject of the subordinate clause in (9a).

- (4) a. *A lion* chased a zebra.
 b. *A zebra* was chased by a lion.
- (5) In principle, he is now capable of carrying out or determining the accuracy of any computation. *Some computations* he may not be able to carry out $\emptyset$ in his head. (Noam Chomsky, 1980, p.221)
- (6) *The woman you were just talking to*, I don't know where *she* went.
- (7) Below the waterfall, a whole mass of enormous glass pipes were dangling down into the river from somewhere high up in the ceiling! *They* really were enormous, *those pipes*. (R. Dahl: *Charlie and the Chocolate Factory*)
- (8) a. Many people believe [that *poinsettias* are poisonous].
 b. Many people believe *poinsettias* to be poisonous.
- (9) a. It seems [that *light energy* will be an important subject of scientific research in the future].
 b. *Light energy* seems to be an important subject of scientific research in the future.

2.1.3 Topic markers

Last but not least, another simple diagnostic to topichood is directly related to the expressions concerning aboutness, such as “speaking of X”, “as for X”, “with regard to X”, “considering X” and “about X”, etc. These words and phrases are felicitous signals denoting X is the topic of the sentence. This is especially the case when these topic markers occur at the beginning of a sentence, which coincides with the preposing principle we have just discussed above.

- (10) **About** *the Rebellion and its results* he would express no opinion. (George Orwell: *Animal Farm*)
- (11) Miss Bingley was engrossed by Mr. Darcy, her sister scarcely less so; and **as for** *Mr. Hurst*, by whom Elizabeth sat, he was an indolent man, who lived only to eat, drink, and play at cards; who, when he found her to prefer a plain dish to a ragout, had nothing to say to her. (Jane Austen: *Pride and Prejudice*)

2.2 Focus

Turning now to **Focus**, I'd like to make a distinction between *Informational focus* and *Contrastive focus* to clarify this general notion. The informational focus of an utterance is the informative or unpredictable part of the assertion that the utterance expresses; it is the information that is not presupposed by the utterance, and rather is attributed to, or predicated of, the topic of that utterance. The typical example of informational focus is the informative part of an answer to a *wh*-question, as *Bill* in (12). By comparison, contrastive focus signals the existence of a set of possible referents that may fill the same position and picks one referent out of that set. It is usually used to lead the addressee's attention to a particular discourse entity, such as a new topic being introduced (topic shift), or one constituent forming contrast with something else. *Helen* in (13a) and *the shirt* in (13b) are examples of purely contrastive focus. What's worth noting here is that, unlike Topic, Focus is not necessarily embodied by a referential expression. In its broad sense, focus refers to all the new information a predicate holds, as shown in (14). However, since the current study is anaphora-oriented, I will center on cases like (12) and (13), where the focused element is nominal in nature and may act as a potential antecedent for anaphora resolution.

- (12) A: Do you know **who** called the meeting?
 B: **Bill** called the meeting. (Gundel, 1993, p.295)
- (13) a. **Helen** bought you the shirt, not Jane.

b. Helen bought you **the shirt**, not the sweater.

- (14) Lee spent the next three months immobilized with a metal halo screwed into his skull to prevent further injury to his neck. What hurt most was that **his doctors didn't talk to him about his prognosis**. (*Reader's Digest*, 2006(8), p.159)

In the identification of focus, we notice a blurring of the distinction between informational focus and contrastive focus because they are coded by similar type of linguistic prominence. To simplify the focus-identifying process, we will not highlight the distinction between the two subcategories unless necessary. There have been many studies justifying the salience weight of focused elements and shedding light on pronoun processing. For instance, Cutler and Fodor (1979) found that detection time for a target phoneme was shorter when the phoneme was in a focused word. Almor (1999) also claimed that antecedents that were in focus appeared to elicit shorter reading time for fuller forms of co-reference, and that such focusing influenced the ease with which coreference was established. Focus is salient because of its newness in relation to the topic of the sentence. Generally speaking, the focus status of an NP is mainly designated by the MARKEDNESS of its prosodic and syntactic properties of the constituent.

2.2.1 Prosodic prominence

The association between prosodic prominence and focus has been shown to hold in a variety of languages, and is widely believed to be universal. In the absence of semantic and pragmatic cues, the focused constituent often carries a **pitch accent**, typically associated with word stress. To simplify the focus-identifying process, we will not highlight the distinction between the two subcategories unless necessary. In written discourse, vocal stress and pitch accent are often simulated through **italics**, **bold type** or **upper case**. Example (15) appears in a print advertisement in *Reader's Digest*, the brand name *Mirapex* easily draws the reader's attention due to the large capital letters. In (16), the author highlights *Magic Formula* through capitalizing its initial letters. Mainly due to the effect of focalization, the pronoun *It* in the following sentence is interpreted as *Magic Formula* rather than other possible antecedents like *attention* or *the main point of your message*.

- (15) Help you restless legs find relief with MIRAPEX. (*Reader's Digest*, 2007(8), p.11)

- (16) By using the Magic Formula you can be certain of gaining attention and focusing it upon the main point of your message. It cautions against indulgence in vapid opening remarks, such as: "I didn't have time to prepare this talk very well," or "When your chairman asked me to talk on this subject, I wondered why he selected me." (Dale Carnegie: *The Quick and Easy Way to Effective Speaking*)

2.2.2 Cleft constructions

In addition to pitch accent, contrastive focus may be syntactically coded by placing the relevant constituent in a syntactically prominent position. The most prototypical and efficient syntactic structure for focus marking is the **Cleft construction**. To be more specific, cleft structures reliably give the clefted constituent a contrastive interpretation. In uttering a cleft sentence, the addressor splits the content of a single proposition into two clauses thereby structurally singling out the clefted constituent and focusing attention on it. There are two main types of Cleft construction: **It-cleft** and **pseudo-cleft**, the former is exemplified by (17), the latter by (18) and (19). (All the clefted constituents are indicated by *italics*.)

- (17) **It** was *legendary SNL creator Lorne Michaels* who recommended him as Letterman's replacement. (*Reader's Digest*, 2006 (9), p.120)

- (18) Out of all the episodes we did, **the one** that really worked was *the one Jeff wrote entirely himself*. (*Reader's Digest*, 2006 (9), p.140)

- (19) Seeing patients exhausted him, both physically and emotionally. **What** now fired him up was *the compulsive sketching he did in between appointments*. (*Reader's Digest*, 2006 (1), p.113)

To sum up, one might wonder why topic and focus, which are considered distinct by theories of information structure, nonetheless appear to function similarly with respect to salience marking. One property that may explain their common effect is that they both presuppose the existence of their referent. Topics, as already discussed above, necessarily presuppose the existence of their referent. Similarly, in order for a cleft construction to be felicitous, elements that are within the scope of focus are presumed to exist prior to the utterance, even if they have not been previously given in the discourse. Further support of this underlying similarity between topic and focus status comes from Kuno (1976, 1987), who essentially argues that the clefted focus in an it-cleft construction must be able to be the topic of the corresponding non-clefted version of the sentence.

Although both **topic** and **focus** serve to increase the cognitive prominence of the linguistically prominent referents in

discourse, it is worth noting that they play slightly different roles in pronoun antecedent selection. On the whole, topic serves as the default antecedent of a pronoun, while focus acts as a marked candidate and carries more salience weight. Because of the bias towards the more prominent antecedent in anaphora resolution, this subtle distinction casts influence on the design of many anaphora resolution algorithms, in which the focused element usually ranks higher than topic on the list of potential antecedents. Such a resolution strategy is justified by Arnold's corpus study (1998), in which he found that the clefted focus element was referred to more often than any other referents within the sentence, and the subject (i.e. topic) was the most likely to be referred back to on the absence of clefted focus.

3. The Formal Bias for Referential Prominence: Pronominalization

There has been some work claiming that the referential form has an effect on salience assessment, but this viewpoint has not received as much attention as other factors. In fact, besides topicality and focalization, a reduced referential form also increases a referent's salience weight and thus makes it more likely to be referred to in upcoming discourse. In other words, a referent will become more salient for future reference if it is expressed with a pronoun rather than a definite NP. This assumption is well-grounded in both linguistic theory and psycholinguistic research (Chafe, 1976; Givón, 1995; Kameyama, 1999, etc.).

Generally speaking, the use of a pronoun signals to the addressee that the addressor is continuing to talk about the same thing and thus keeping it in the spotlight. From the viewpoint of coherence construction, this explains why pronouns are frequently used in a topic chain. Example (20) illustrates the most typical occurring type of pronouns as a continuation of topic. Being the central topic of the discourse segment, *Mr. Pilkington* has been referred to repeatedly with pronominal forms like *his* in (20a) and *he* in (20b, c).

(20) a. **Mr. Pilkington**, of Foxwood, had stood up, *his* mug in *his* hand.

b. In a moment, *he* said, *he* would ask the present company to drink a toast.

c. But before doing so there were a few words that *he* felt it incumbent upon him to say.

(George Orwell: *Animal Farm*)

More importantly, such a saliency bias for pronouns sheds light on anaphora resolution when there is referential ambiguity. For instance, in the interpretation of the pronoun *she* in (21c), we find two antecedent candidates in (21b): *Miss Bingley* and *Elizabeth (her)*. Although *Miss Bingley* appears in the subject position, *Elizabeth* outweighs *Miss Bingley* in saliency for the sake of pronominalization. Such a referential choice is further justified if we take the whole discourse segment into account. As a matter of fact, *Elizabeth* functions as the Global Topic of (21), with its centering status fully verified through the use of following pronouns: *her* in (21b) and (21d), *she* in (21c).

(21) a. When the clock struck three, **Elizabeth** felt that she must go, and very unwillingly said so.

b. Miss Bingley offered her (=Elizabeth) the carriage,

c. and she only wanted a little pressing to accept it,

d. when Jane testified such concern in parting with her,

e. that Miss Bingley was obliged to convert the offer of the chaise to an invitation to remain at Netherfield for the present.

(Jane Austen: *Pride and Prejudice*)

In brief, being pronominalized makes a referent salient in the future. It should be noted that this constraint is optional. That is, if there are no pronouns in the set of potential antecedents, this salience-form constraint will not have any effect in the final preference over candidate interpretation.

4. The Viewpoint Effect on Salience Denotation: Empathy

In its broad sense, **Empathy** refers to the addressor's adoption of a perspective or standpoint in describing an event or state. In fact, the addressor may relate information from various "camera angles". For instance, he can place himself at a distance from the participants of the event/state and give an objective, detached description. Alternatively, he can position himself closer to one participant than to the others, or in some special cases, he can completely identify himself with one participant and describe the event/state from this participant's camera angle.

In Kuno's functional approach to the speaker's identification, empathy is defined as a gradient notion. To be more specific, the degree of the speaker's empathy with *x*, **E(x)**, ranges from 0 to 1, with $E(x) = 1$ signifying his/her total identification with *x* and $E(x) = 0$ signifying a total lack of identification. Accordingly, I would like to denote the two ends of this continuum as **subjective point of view** and **objective point of view** respectively. The former reflects the

consciousness of addressor through presenting his perception, evaluations, judgments, or inner thoughts; the latter is featured by its independence of any character's consciousness.

With respect to anaphora processing, empathy plays an influential role in salience ranking. It has generally been assumed that the contribution of point of view is entirely separate from the syntactic constraints that determine the core anaphora facts. Many scholars (Kuno & Kaburaki, 1977; Kameyama, 1985; Walker, Iida, & Cote 1994) first notice its predictive power of salience weight in the interpretation of zero pronoun in Japanese, where each predicate selects one of its arguments as the locus of empathy. Based on his analyses of Turkish discourse, Turan (1995) argues that empathy is relevant to Western languages as well. In English, the representation and interpretation of discourse entities in attentional state are also sensitive to the addressor's choice of viewpoint. This is particularly the case in a discourse segment with a subjective point of view, where the salience weight of a discourse entity partly depends on how it is perceived by the subject of consciousness. In (22), for example, Elizabeth is salient as the subject of consciousness. The empathic prominence is indicated by the phrase in Elizabeth's mind, and it affects the attentional center indeed. This explains why the pronoun *she* in (22d) refers back to *Elizabeth* in (22a) rather than a nearer potential antecedent, *Mrs. Reynolds* in (22b).

- (22) a. There was certainly at this moment, **in Elizabeth's mind**, a more gentle sensation towards the original than she had ever felt at the height of their acquaintance.
 b. The commendation bestowed on him by Mrs. Reynolds was of no trifling nature.
 c. What praise is more valuable than the praise of an intelligent servant?
 d. As a brother, a landlord, a master, **she** considered how many people's happiness were in his guardianship!

(Jane Austen: *Pride and Prejudice*)

To detect the projector of the limelight, i.e. the subject of consciousness in automatic processing, we need the help of certain **empathy-marking verbs** whose argument structure corresponds to a particular point of view of the event. Typically, verbs denoting perception like *see*, *look at*, *feel*, *appear* and verbs representing inner thought like *remember*, *interest*, *consider* are empathetically deictic. It is always the direct object of this kind of verb that will gain more attention and thus become more salient. In addition, there are expressions that refer to a character's point of view, such as *The thought cross her mind*. With such expressions, it is the experiencer, which is often in object position, rather than the grammatical subject, that should be ranked higher. Consider the potentially ambiguous pronoun *her* in (23b), which might be expected to realize the entities in (23a), i.e., *Mrs. Bennet* or *her eldest daughter*. Although *Mrs. Bennet* is more salient in terms of grammatical function, the definite pronoun *her* in (23b) co-refers with the object *her eldest daughter* because it follows the perceptual verb *seen* and thus gains more empathetic prominence.

- (23) a. Mrs. Bennet; had **seen** her eldest daughter; much admired by the Netherfield party.
 b. Mr. Bingley had danced with her; twice,
 c. and she; had been distinguished by his sisters.

(Jane Austen: *Pride and Prejudice*)

5. Conclusion

On the whole, the discourse entity that is at the center of attention will be assigned the highest saliency and tend to be referred to with the most reduced anaphoric form. Motivated by such a cross-linguistic consensus, this paper tries to identify the attentional center through assessing the relative salience of different discourse entities. Three linguistic properties that have been claimed to influence the salience weight in English are examined in turn: **information status**, **referential form**, and **empathy**. It is hoped that the salience assessment outlined here will contribute to more adequate and insightful analyses of anaphora processing. It's worth noting that the discourse samples are primarily drawn from literary works and magazine articles, the exploration of saliency in spoken monologues and the intricacies of dynamic conversations are left open for future investigation.

References

- Almor, A. (1999). Noun-phrase anaphora and focus: the informational load hypothesis. *Psychological Review*, 106(4), 748-765. <http://dx.doi.org/10.1037/0033-295X.106.4.748>
- Arnold, J. (1998). Reference form and discourse patterns. Doctoral Dissertation. Stanford University.
- Ariel, M. (1990). *Accessing Noun-phrase Antecedents*. London: Routledge.
- Chafe, W. L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Li C (Ed.) *Subject and Topic*. New York: Academic Press. pp. 25-56.
- Chomsky, N. (1980). *Rules and Representations*. Oxford: Basil Blackwell.
- Crawley, R. A., Stevenson, R. J., & Kleinman, D. (1990). The use of heuristic strategies in the interpretation of pronouns. *Journal of Psycholinguistic Research*, 19, 245-264. <http://dx.doi.org/10.1007/BF01077259>
- Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, 7 (1), 49-59. [http://dx.doi.org/10.1016/0010-0277\(79\)90010-6](http://dx.doi.org/10.1016/0010-0277(79)90010-6)
- Firbas, J. (1966). Non-thematic subjects in contemporary English. *Travaux Linguistiques de Prague*, 2, 239-56.
- Firbas, J. (1992). *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: Cambridge University Press.
- Givón, T. (Ed.). (1983). *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam; Philadelphia: J. Benjamin's Pub.
- Gundel, J., Hedberg N., & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, 69, 274-307. <http://dx.doi.org/10.2307/416535>
- Givón, T. (1995). *Functionism and Grammar*. Amsterdam: John Benjamins Publishing Company.
- Gundel, J., Hedberg N., & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, 69, 274-307. <http://dx.doi.org/10.2307/416535>
- Kameyama, M. (1985). Zero anaphora: the case of Japanese. Doctoral dissertation. Stanford University.
- Kameyama, M. (1999). Stressed and unstressed pronouns: complementary preferences. In Peter Bosch and Rob van der Sandt (Eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge: Cambridge University Press. pp. 306-321.
- Kuno, S., & Kaburaki, E. (1977). Empathy and Syntax. *Linguistic Inquiry*, 8, 627-72.
- Kuno, S. (1976). Subject, Theme, and the speaker's empathy: a reexamination of relativization phenomena. In C. Li (Ed) *Subject and Topic*. New York: Academic Press. pp. 417-444.
- Kuno, S. (1987). *Functional Syntax: Anaphora, Discourse and Empathy*. Chicago: University of Chicago Press.
- Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- Marslen-Wilson, W., Tyler, L. K., & Koster, C. (1993). Integrative processes in utterance resolution. *Mem Lang*, 32, 647-666. <http://dx.doi.org/10.1006/jmla.1993.1033>
- Reinhart, T. (1982). Pragmatics and linguistics: an analysis of sentence topics. *Philosophica*, 27, 53-94.
- Stevenson R. J., Nelson, A. W. R., & Stenning, K. 1995. The role of parallelism in strategies of pronoun comprehension. *Lang Speech*, 38(4), 393-418.
- Turan, U. D. (1995). Null vs. Overt Subjects in Turkish Discourse: a centering analysis. Doctoral dissertation. University of Pennsylvania.
- Walker, M. A., Iida, M., & Cote, S. (1994). Japanese Discourse and the Process of Centering, *Computational Linguistics*, 20(2), 193-232.

Notes:

1. Although other constituent types may also be perceived as topics, the most clearly observable and objectively definable linguistic features of topics are those associated with NP's.
2. When discussing clausal constituency in his functional grammar system, Halliday distinguished three broad definitions of Subject:
 - a. Psychological subject: that which is the concern of the message;
 - b. Grammatical subject: that of which something is predicated;
 - c. Logical subject: the doer of the action.

According to him, these three concepts are not different varieties of a general notion Subject. Instead, they are three separate and distinct functions, relabeled as "Theme", "Subject" and "Actor" by Halliday respectively. It's worth noting that the term subject is used here in the sense of Halliday's Grammatical subject if not clearly designated.

3. It's worth mentioning that informational focus and contrastive focus are associated with distinct pitch accents in English. Informational focus is coded by what Bolinger (1961) and Jackendoff (1972) call an **A accent**, while a contrastive focus is typically marked by what they call a **B accent**. As the current study focus on written texts, I will not further discuss the distinction between the two prosodic strategies here.