Identifier 'So' and the Information Status of Discourse Referents

Andrew Kehler — SRI International[*]

Gregory Ward — Northwestern University[†]

LSA Annual Meeting
Los Angeles, CA
January 8, 1999

1 Reference and the Information Status of Referents

A speaker's choice of referring expression must be conditioned on the *information status* of the referent

1.1 A Unidimensional Account: Gundel et al., 1993

Propose a *Givenness Hierarchy* relating information status to linguistic form:

in focus	>	activated	>	familiar	>	uniquely identifiable	>	referential	>	type identifiable
it		*that*		*that* N *this* *this* N		*the* N		indef *this* N		*a* N

1.2 A Bidimensional Account: Prince, 1992

Distinguishes *hearer status* from *discourse status*:

- Hearer Old/New: whether or not the referent is known to the hearer
- Discourse Old/New: whether or not the referent is already evoked

[*]Artificial Intelligence Center, 333 Ravenswood Avenue, Menlo Park, CA 94025, kehler@ai.sri.com.
[†]Department of Linguistics, 2016 Sheridan Road, Evanston, IL 60208, gw@nwu.edu.

1.3 Linguistic vs. Situational Evocation

Neither account distinguishes between linguistically- and situationally-evoked referents, i.e., *he* is categorized the same way in (1) and (2).

(1) Bill Clinton entered the press room. *He* is about to give a speech.

(2) [Bill Clinton enters the press room] *He* is about to give a speech.

Forms which require linguistic evocation were classified by Sag and Hankamer (1984) as *Ellipsis*:

- e.g., gapping, stripping, VP-ellipsis
- antecedents are syntactic logical forms, implying that they may not be situationally-evoked (i.e., 'pragmatically-controlled')
- not within the scope of theories of information status concerned only with semantically-based reference (i.e., *Model Interpretive Anaphora*)

In this talk, we discuss reference with 'identifier *so*', and

- show it is semantically resolved yet does not allow for situationally-evoked referents,
- argue that it is in the purview of theories of information status but not categorizable within them, and
- propose a modification to Prince's framework to handle it

2 The Properties of *So* Anaphora

'Identifier *so*' (Bolinger, 1972) is used to establish reference to a contextually salient event of the type denoted by the verb it modifies

(3) "...and with complete premeditation resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system." He was *so strangled* on Aug. 26, 1975, in his bed most cruelly. (*Chicago Tribune* 12/15/94)

(4) In fact, in substantiating these fears, Judge Bork again essentially concedes that economic freedom is a component of the Constitution: "We already have clauses that could be used to protect economic freedom – and were *so used*." (*Wall Street Journal*)

Thus we are not considering the numerous other uses of *so*:

- Intensifier *so*: There are *so* many uses of 'so'!
- 'Also' *so*: Bill gave a speech and *so* did Hillary.
- Generation X *so*: I'm like, *so* going out with him.
- and *so* forth...

Preverbal and postverbal uses of identifier *so* have different properties; we focus on the preverbal form exemplified in (3) and (4)

Preverbal *so* requires a hearer-old referent:

(5) His Imperial Majesty Haile Selassie was head of the feudal system.

a. #He was so strangled on Aug. 26, 1975, in his bed most cruelly.

b. He was strangled on Aug. 26, 1975, in his bed most cruelly.

(6) In fact, in substantiating these fears, Judge Bork again essentially concedes that economic freedom is a component of the Constitution:

a. #"We already have clauses that could be so used by opponents of the administration's trade policy."

b "We already have clauses that could be used by opponents of the administration's trade policy."

However, the referent *cannot* be situationally-evoked:

(7) [A and B together have just witnessed Haile Selassie being murdered by strangulation]

a. A: #He was so strangled most cruelly.

b. A: He was strangled most cruelly.

(8) [A and B have just witnessed a vote in Congress that repealed an amendment of the Constitution]

a. A: #By so changing the constitution, Congress is setting a dangerous precedent.

b. A: By changing the constitution, Congress is setting a dangerous precedent.

The referent must be salient:

(9) With complete premeditation, they resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system. They also resolved to commit a variety of other violent acts, although those would come somewhat later. Obviously, these people were very prone to violence.

a. #Selassie was so strangled on Aug. 26, 1975, in his bed most cruelly.

b. Selassie was strangled on Aug. 26, 1975, in his bed most cruelly.

Can be used when the referent must be inferred:

(10) Only about \$20 billion of Latin America's \$350 billion foreign bank debt is going through debt-reduction programs this year. But that's about four times the amount *so treated* last year. (*Wall Street Journal*)

(11) Regarding a possible Elvis Presley stamp, Postmaster General Frank notes that anyone *so honored* must be "demonstrably dead" for 10 years. (*Wall Street Journal*)

Previous examples suggest an analysis similar to *(in) that way*, but:

(12) "...and with complete premeditation resolved that His Imperial Majesty Haile Selassie should be strangled because he was head of the feudal system." He was (√ so strangled)/(# strangled in that way) on Aug. 26, 1975, in his bed most cruelly.

(13) [A and B have just witnessed Haile Selassie being murdered by strangulation]

a. A: I can't believe he was strangled (in) that way!

b. A: #I can't believe he was so strangled!

Unlike *so*, *(in) that way* can establish reference to a manner modifying a distinct event previously evoked in the discourse:

(14) A: John sings beautifully.

B: Yeah, and he writes that way too.

B': #Yeah, and he so writes too.

In summary, the referent of *so+V* is:

- hearer-old and salient
- semantically-resolved, without requiring a 'syntactic' antecedent
- linguistically-evoked

These same properties hold of *do so* (cf. *so doing*) (Kehler and Ward, 1995)

Does not require a syntactic antecedent:

(15) Section 1 provides the examples to be derived by Gapping, and a formulation of Gapping capable of doing so. [= deriving the examples] (text of Neijt (1981))

(16) There was a lot more negativity to dwell on, if anyone wished to do so. [= dwell on more negativity]

(17) With or without the celebration, Belcourt is well worth seeing, and you can do so year round. [= see Belcourt]

'Do so' nonetheless fails for situationally-evoked referents:

(18) [A and B have just witnessed Haile Selassie being murdered by strangulation]

A: # I can't believe they did so!

Therefore,

- theories must allow for referential forms which are resolved semantically but do not allow situationally-evoked referents (Kehler and Ward, 1995)
- as a legitimate form of anaphora, preverbal *so* falls within the purview of theories of information status

3 Ramifications for Theories of Information Status

Unidimensional theories do not offer a place for distinguishing between forms requiring *in focus* referents that allow situationally-evoked referents (e.g., pronouns), and those which do not (e.g., *so* anaphora)

In treating both linguistically- and situationally-evoked referents as discourse-old, Prince's system similarly does not allow for this distinction

Prince says:

> In fact, it is interesting that in English, at least, there is virtually no marking of an NP with respect to the Discourse-status of the entity it represents. Of course, if an NP is indefinite and is thereby understood as evoking something Hearer-New, we can infer Discourse-new. However, if it is not so marked, then, with one exception, we cannot tell from its form whether it has occurred before in the discourse. (Prince, 1992, p. 304)

Exception is the case of pronouns:

> Pronouns indicate that the entities they represent are salient, i.e. appropriately in the hearer's consciousness ... at that point in the construction of the discourse model. Therefore, they are presumably already in the discourse model. Therefore, they are Discourse-old. However, at any point in (discourse) time, only a subset, usually proper, of the entities already evoked are salient and hence representable by a pronoun. (Prince, 1992, p. 304)

Use **Salient** to refer to highly activated or accessible referents, regardless of type of evocation

Use **Discourse old** to encompass only referents which are *explicitly* introduced into the discourse, i.e., linguistically-evoked

Thus:

- Pronouns require a salient referent
- Identifier *so* requires a salient, discourse-old referent

4 Conclusion

- *So* anaphora imposes a combination of constraints on the information status of its referent not found for any NP anaphor in English: hearer-old, salient, and linguistically-evoked
- *So* anaphora thus requires that theories of information status be revised
- Prince's bidimensional system can accommodate *so* anaphora by distinguishing a third dimension, *salience*, and constraining discourse-oldness to encompass only linguistically-evoked referents

Acknowledgements

This research was supported by National Science Foundation Grant IIS-9619126 (Kehler), and NIDCD grant R01-DC01240 (Ward).

References

Bolinger, Dwight. 1972. *Degree Words*. Mouton, The Hague, Paris.

Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307, June.

Kehler, Andrew and Gregory Ward. 1995. On the anaphoric status of *do so*. Paper presented at the Annual Meeting of the Linguistic Society of America, New Orleans, January.

Prince, Ellen. 1992. The ZPG letter: Subjects, definiteness, and information-status. In Sandra Thompson and William Mann, editors, *Discourse Description: Diverse Analyses of a Fundraising Text*. John Benjamins B.V., Philadelphia/Amsterdam, pages 295–325.

Sag, Ivan and Jorge Hankamer. 1984. Toward a theory of anaphoric processing. *Linguistics and Philosophy*, 7:325–345.

Example

(19) [A and B are talking about Lisa Marie Presley, when Oprah Winfrey walks onto the subway]

A to B: Apparently *she*$_1$ (=Winfrey) is a very smart businesswoman. *A local newspaper* and *People Magazine* each recently listed the richest women in the world, and in each case *she*$_2$ was one of the women *so listed*. Actually, for that matter, *Lisa Marie* was too.

	Discourse New	Discourse Old
Hearer New	**a local newspaper**	[not applicable]
Hearer Old (Not Salient)	**People Magazine**	**Lisa Marie**
Hearer Old (Salient)	**she**$_1$	**she**$_2$ **so listed**