



## Deseret Language and Linguistic Society Symposium

---

Volume 25 | Issue 1

Article 10

---

2-19-1999

### Who is "him"? Determining Pronominal Reference

Deryle Lonsdale

Follow this and additional works at: <https://scholarsarchive.byu.edu/dlls>

---

#### BYU ScholarsArchive Citation

Lonsdale, Deryle (1999) "Who is "him"? Determining Pronominal Reference," *Deseret Language and Linguistic Society Symposium*: Vol. 25 : Iss. 1 , Article 10.

Available at: <https://scholarsarchive.byu.edu/dlls/vol25/iss1/10>

This Article is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Deseret Language and Linguistic Society Symposium by an authorized editor of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

# Who is “him”? Determining Pronominal Reference

Deryle Lonsdale

## 1. Abstract

In this paper we look at contexts where determining the references of pronouns is problematic to varying degrees. A computational technique, discourse centering, is applied to those contexts and predictions about preferred scenarios are discussed. Though no comprehensive solution is proposed, this informal presentation highlights places in the scriptures (particularly in the Book of Mormon) where such situations arise.

## 2. Introduction

In this paper we discuss a technique for determining (or picking) the appropriate reference from among (a set of) possible referents. Of interest to this conference is that the data to be used in this presentation, by way of illustration, is taken from the scriptures.

Computational techniques have been developed in recent years to encode algorithms for determining discourse reference which have been explored in the syntactic and pragmatic literature. One such technique, discourse centering [Gordon et al., 1993, Brennan, 1995, Grosz et al., 1995], has been applied to some degree of success in a variety of languages [Turan, 1995] and text types.

In Japanese, for example, it is possible to encounter sequences such as the following [Walker et al., 1992]:

(1) Taroo ga kooen o  
Taroo SUBJ park in

sanpositeimasita.  
walking-was

Taroo was taking a walk in the park.

(2) Ziroo ga Ø hunsai no mae  
Ziroo SUBJ OBJ fountain of front  
de mitukemasita.  
in found

Ziroo found (Taroo) in front of the fountain.

(3) Ø Ø kinoo no siai no  
SUBJ OBJ yesterday of game of  
kekka o kikimasita.  
scores OBJ asked

(Ziroo) asked (Taroo) the score of yesterday's game.

In the first sentence, an individual, Taroo, is introduced into the discourse. In the second sentence, a zero pronoun (which is licit in Japanese) is used to refer to the object. Here the “missing” object is taken to refer to the individual introduced in the opening sentence. In the third sentence, two zero pronouns are used -- one for the subject and one for the object. Still, it is possible to understand the underlying meaning, given established syntactic and discourse principles of the Japanese language.

Note that an alternative sequence, where the subject is dropped in the second sentence, is also possible:

- (4) Taroo ga kooen o  
 Taroo SUBJ park in  
  
 sanpositeimasita.  
 walking-was

Taroo was taking a walk in the park.

- (5) Ø Ziroo o hunsai no mae  
 SUBJ Ziroo OBJ fountain of front  
  
 de mitukemasita.  
 in found

(Taroo) found Ziroo in front of the fountain.

- (6) Ø Ø kinoo no siai no  
 SUBJ OBJ yesterday of game of  
  
 kekka o kikimasita.  
 scores OBJ asked

(Taroo) asked (Ziroo) the score of yesterday's game.

Here, the first sentence is as before, in the second sentence the subject is dropped (or gapped), and (consequently) the third sentence is completely different in its interpretation compared to the previous scenario.

### 3. Fundamental definitions

Implementation of the centering algorithm can be summarized as a three-step process [Walker et al., 1992]:

- GENERATE a set of possible referents
- FILTER out unlikely referents
- RANK the remaining referents

The first step, that of generating referents, is simply a listing of all referents which could possibly be involved in a given instance of anaphor. The second step, filtering, involves the use of standard grammatical procedures to rule out otherwise possible referents. The third step, ranking of plausible referents, involves a procedure called transition ordering. In the next section we will consider the transition ordering step, which is the most technical part of the algorithm.

#### 3.1 Constraints

For each utterance  $U_i$  in a discourse segment  $U_1, \dots, U_m$ :

1. There is precisely one backward looking center  $C_b$ .
2. Every element of the forward centers list,  $C_f(U_i)$ , must be realized in  $U_i$ .
3. The center,  $C_b(U_i)$ , is the highest-ranked element of  $C_f(U_{i-1})$  that is realized in  $U_i$ .

In the next section we will illustrate how these constraints are applied.

#### 3.2 Rules

For each  $U_i$ , in a discourse segment  $U_1, \dots, U_m$ :

1. If some element of  $C_f(U_{i-1})$  is realized as a pronoun in  $U_i$ , then so is  $C_b(U_i)$ .
2. Transition states are ordered:  
 CONTINUE > RETAIN > SMOOTH-SHIFT > SHIFT.

The latter rule states that CONTINUE is a more preferable transition to RETAIN, which is preferred over SMOOTH-SHIFT, and so on.

Table 1 gives a summary of the scenarios leading to various transition values between subsequent utterances.

	$Cb(U_i) = Cb(U_{i-1})$	$Cb(U_i) \neq Cb(U_{i-1})$
$Cb(U_i) = Cp(U_{i-1})$	CONTINUE	SMOOTH-SHIFT
$Cb(U_i) \neq Cp(U_{i-1})$	RETAIN	SHIFT

Table 1: Summary of transition types [Walker et al., 1992].

#### 4. An example

Having sketched the algorithm, we are ready to apply it. Consider the following example, where narrative discourse is taken from the Book of Mormon.

*. . . and Laman went in unto the house of Laban, and he talked with him as he sat in his house. And he desired of Laban the record . . .* (1 Nephi 3:11-12)

In this passage we have three clauses which admit several possible readings, depending on the referents we ascribe to the different pronouns. We next describe, clause-wise, the possible readings and how one can apply the centering algorithm in their analysis.

*Laman<sub>j</sub> went in unto the house of Laban<sub>j</sub>*

Here we have the introduction of two discourse referents which will be tracked over the next several clauses: *Laman* and *Laban*.<sup>1</sup> We index the referents in their order

of appearance, with subscripts *i* and *j* respectively.

*and he talked with him*

Here we have two pronouns, the subject *he* and the oblique *him*, whose reference must be determined. We assume that the subject pronoun could apply to either *Laman* or *Laban*. Given this, though, we are constrained in our interpretation of the pronoun *him*. In particular the latter, by general principles of pronominal and anaphoric binding, cannot likewise refer to the same referent as the subject; disjoint reference is mandatory. Note that:

*John talked with him.*

can only be interpreted with disjoint reference; otherwise one would need to say instead:

*John talked with himself.*

This is a straightforward application of Binding Theory in syntax. The process of ruling out possible indexing scenarios on standard grammatical principles we call “contraindexing.”

Hence, as we consider the present clause we are left with two possible indexings:

. . . *and he<sub>i</sub> talked with him<sub>j</sub>*, or  
 . . . *and he<sub>j</sub> talked with him<sub>i</sub>* where  
 the former can be rephrased  
 . . . *and Laman talked with Laban*,  
 and the latter  
 . . . *and Laban talked with Laman*

Given that there are two possibilities at this point, we will later call on the centering algorithm to rank them.

*as he sat in his house*

In this sentence we have several possible indexings. The two pronominals, *he* and *his*, can each take either *I* or *j* as indices:

*as he<sub>i</sub> sat in his<sub>i</sub> house*  
*he<sub>i</sub> sat in his<sub>j</sub> house*  
*as he<sub>j</sub> sat in his<sub>j</sub> house*  
*as he<sub>j</sub> sat in his<sub>i</sub> house*

meaning, respectively,

*as Laman sat in Laman's house*  
*as Laman sat in Laban's house*  
*as Laban sat in Laman's house*  
*as Laban sat in Laban's house*

Table 2 contains a summary of the plausible indexings for the latter two clauses. The first column, when filled with an asterisk, indicates that the given indexing violates either conindexing or the assumption that the discussion took place in Laban’s house.

	<i>he</i>	<i>talked to him</i>	<i>while he</i>	<i>was in his house</i>	
*	i	i	i	i	conindexing
*	i	i	i	j	conindexing
*	i	i	j	i	conindexing
*	i	j	i	i	Laban’s house
*	j	i	i	i	Laban’s house
	j	i	i	j	
*	i	j	j	i	Laban’s house
*	j	j	i	i	conindexing
*	i	i	j	j	conindexing
*	i	i	i	j	conindexing
*	j	j	j	i	conindexing
*	j	j	j	j	conindexing
*	j	j	j	j	conindexing
*	j	i	j	i	Laban’s house
*	j	j	i	j	conindexing
	i	i	j	i	conindexing
	j	i	j	j	
	i	j	i	j	
	i	j	j	j	

Table 2: Listing of plausible referents for a sentence.

Note that four possible readings still remain for the sentence:

*he spoke with him as he sat in his house*

namely:

*Laban spoke with Laman as Laman sat in Laban’s house*

*Laban spoke with Laman as Laban sat in Laban’s house*

*Laman spoke with Laban as Laman sat in Laban’s house*

*Laman spoke with Laban as Laban sat in Laban’s house*

respectively.

It should be noted that, exceptionally in this case, the set of those four possible readings is not problematic: with any one reading we still have a discussion between *Laman* and *Laban* and which takes place at the latter’s house. The distinction drawn in

these alternatives is with whose actions are being discussed. One could imagine other passages of discourse where such situations would be less than straightforward, and where complete disambiguation might be necessary for purposes of clarity.

Now given these four possibilities, it is possible to assign them ranking given the centering algorithm. We will trace through an application of the algorithm to the three sentences in question to illustrate how this is done.

First, consider the sentence:

*Laman<sub>i</sub> went in unto the house of  
Laban<sub>j</sub>*

Here the backward-looking center Cb is the entity *Laman* since the discourse previous to the clause in question focuses on him. The set of forward-looking centers, Cf, contains both *Laman* and *Laban* since both are mentioned in this sentence. We will take Cb(1) and Cf(1) to mean Cb and Cf (respectively) for the first sentence.

$Cb(1) = (Laman)$   
 $Cf(1) = (Laman, Laban)$

Note that in this case *Laman* is both Cp(1)<sup>2</sup> and Cb(1).

We next consider the sentence:

*and he talked with him*

Here we have two possible licit indexings:

*he<sub>i</sub> talked with him<sub>j</sub>* (call this sentence 2a)  
*he<sub>j</sub> talked with him<sub>i</sub>* (call this sentence 2b)

Consider first the former sentence, 2a. We set Cb to *Laman* since this is the

most prominent realized member of Cf from the previous utterance (i.e. the value Cp). We map our assumed indexings for this sentence into Cf.

$Cb(2a) = \{Laman\}$   
 $Cf(2b) = \{Laban, Laman\}$

Here we also have a CONTINUE transition from sentence 1 to sentence 2b since Cb(2b) = Cp(1) and Cb(2b) = (Cb)1: in this case all values again refer to *Laman*.

Next, we consider the third clause:

*as he sat in his house*

Here, even though we assume *his house* refers to *Laban's*, we have two possible readings, depending on the value for *he*:

*as Laman sat in Laban's house* (call this sentence 3a)  
*as Laban sat in Laban's house* (call this sentence 3b)

$Cf(3a) = \{Laman, Laban\}$   
 $Cf(3b) = \{Laban, Laban\}$

In considering how this clause combines with the previous one, we now get four possible scenarios for consideration:

- If we follow the first reading for the previous clause (2a) and assume the first reading for this one (3a), we obtain a backward looking center value Cb(2a+3a) = {*Laman*} since this is the most prominent member of Cf(2a) realized in 3a. Consequently, we have a CONTINUE transition since Cb(2a+3a) = Cp(2a) and

$Cb(2a+3a) = Cb(2a)$ , all values yielding *Laman*.

- However, if we follow the first reading for the previous clause (2a) and assume the second reading for the present clause (3b), we obtain  $Cb(2a+3b) = \{Laban\}$  since *Laban* is the most prominent member of  $Cf(2a)$  realized in 3b. Hence in this case we have a SHIFT transition since  $Cb(2a+3b) \neq Cp(2a)$  and  $Cb(2a+3b) \neq Cb(2a)$ .
- If we follow the second reading (2b) for the previous clause and assume the first reading for this one, we obtain for this clause  $Cb(2b+3a) =$  and  $Cf(3a) = \{Laman, Laban\}$ . This is a SMOOTH-SHIFT transition since  $Cb(3a) = Cp(2b)$  but  $Cb(3a) \neq Cb(2b)$ .
- If we follow the second reading for both the previous clause (2b) and this one (3b), we obtain for this clause  $Cb(2b + 3b) =$

$\{Laban\}$  and  $Cf = \{Laban, Laban\}$ . This, too, is a SMOOTH-SHIFT transition following reasoning sketched in the previous item.

Taking stock of our progress to this stage, we can summarize the transition patterns in terms of the possible indexing scenarios we have adopted:

i j i j CONTINUE + CONTINUE  
 i j j j CONTINUE + SHIFT  
 j i i j CONTINUE + SMOOTH-SHIFT  
 j i j j CONTINUE + SMOOTH-SHIFT

Consider the next clause:

*And he desired of Laban the records . . .*

Here, contraindexing rules out *Laban* as a possible antecedent for the pronoun, so *Laman* obviously is the referent here. Given his fact, one more iteration of the centering algorithm (which we leave as an exercise to the reader) results in the following transition sequences:

i j i j	CONTINUE +CONTINUE + CONTINUE	1
i j j j	CONTINUE + SHIFT + CONTINUE	3
j i i j	CONTINUE + SMOOTH-SHIFT + SMOOTH-SHIFT	4
j i j j	CONTINUE-SMOOTH-SHIFT + CONTINUE	2

The numbers in the last column represent preference rankings based on the transition hierarchy. Hence the algorithm judges that the most natural, or least jarring, sequence would be paraphrased as:

*. . . and Laman went in unto the house of Laban, and Laban talked with Laman as Laman sat in Laban's house. And Laban desired of Laban the records . . .*

A schematic illustration of application of the centering algorithm for the

steps we have considered is given in Figure 1.

**5. Other examples**

Appendixes A and B list several passages from the (first half of the) Book of Mormon and from the Pentateuch (King James Version) respectively which contain passages ambiguous in the ways our example passage was.

Many involve dialogue; in these cases contextual clues from the conversational turns can inform the reader as to proper referent identification. In some cases, though, possibilities abound;

occasionally such passages require significant deliberate attention to the resolution process.

*And Joseph . . . went up to meet Israel his father . . . and presented himself unto him; and he fell on his neck, and wept on his neck a good while (Genesis 46:29).*

One might wonder, in this passage, who fell on whose neck.

By way of illustration we give two other similar passages, one from the Bible and another from the Book of Mormon.

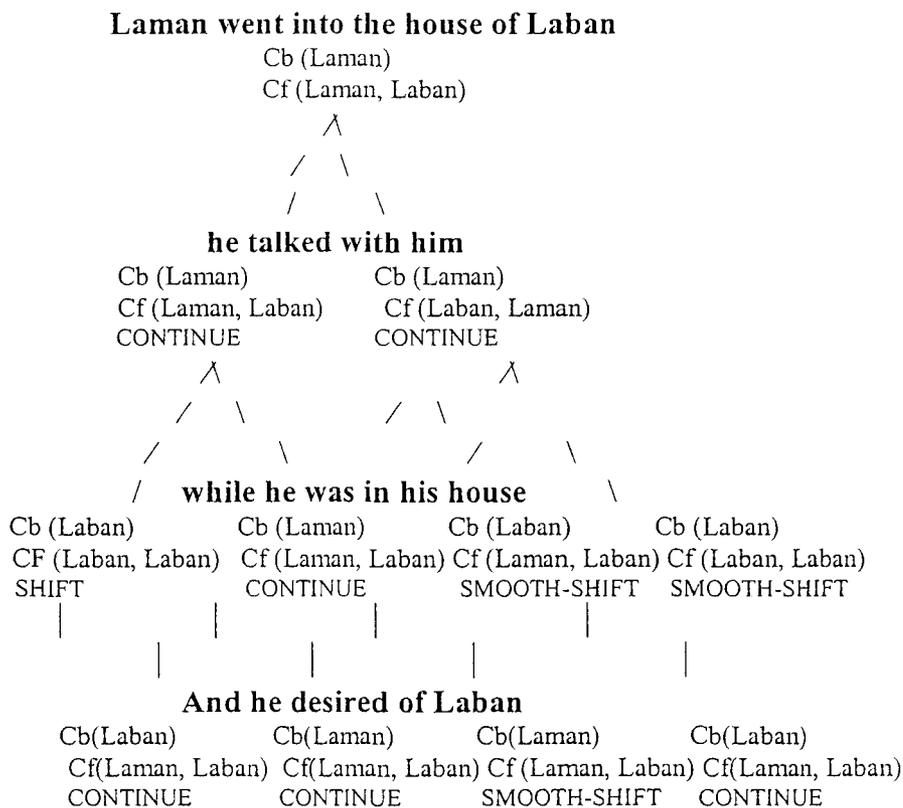


Figure 1: Decision tree for centering-based processing.

*And if a man smite his servant . . . and he die under his hand; he shall be surely punished. Notwithstanding, if he continue a day or two, he shall not be punished: for he is his money (Exodus 21:19-20).*

*Now Alma, seeing that the words of Amulek had silenced Zeezrom, for he beheld that Amulek had caught him in his lying and deceiving to destroy him, and seeing that he began to speak unto him, and to establish the words of Amulek . . . (Alma 12:1).*

The latter passage is so massively ambiguous that concerted effort must be made to completely resolve referents. Hand-application of the centering algorithm for this example becomes quickly impossible.

## 6. Conclusion

So far we have discussed a previously proposed algorithm for discourse centering, and shown its application to sample passages of scripture. In this section we pause to mention the implications for such an approach and possible further related research.

First and foremost, the centering approach was developed as a method for calculating and ranking preferences in ambiguous environments. As such it is not just a subjective system but embodies a quantitative method for ranking preferred readings in complex passages. This is useful in the context of natural language processing, since the algorithm can be coded up in a computer language and used by a computer in text or dialogue understanding.

Secondly, the approach is claimed to provide a predictive framework for

processing. This implies that one could, for example, investigate on-line processing in human subjects by interrupting their reading of certain passages and asking what interpretations are preferred at that time, and how such preferences might change as more text is encountered.

An intriguing area for further research is how these mechanism might interface with grammatical principles. It has been noted that some languages process lexical, morphologic, or syntactic devices that are used to signal transitions in discourse reference. Both formalist and functionalist literatures discuss various methods that different languages use for this purpose.<sup>3</sup>

More speculatively, assuming the availability of grammatical devices for marking discourse referent transition in a given source language, they could perhaps be observed in a target language translation which adheres closely to the source in content and style. If particular devices used to flag infelicitous, jarring transitions in a source language were rendered directly via translation, their reflects would be discernable in the target language text. Occasional mention has been made of such frequently occurring discourse transition markers as “*Now . . .*,” and so on, as they appear in English translations of scripture. Could there be a discourse-transition component in their appearance and frequent use?

As a descriptive tool, evaluation metric, and springboard for further study, the centering algorithm provides interesting insights into discourse referent tracking. Though we have only touched on a few of its basic properties, it promises to be helpful for those interested in studying how

languages use referents and how we interpret them.

### End notes

<sup>1</sup>In fact, there is also a third, *the house*, which will not be treated in our discussion.

<sup>2</sup>Cp is defined as the first, or most prominent, member of Cf.

<sup>3</sup>For example, see [Bates, 1997].

### References

- [Bates, 1997] Bates, D. (1997). Semantic roles and referent tracking in Martha Lamont's 'Pheasant and Raven'. In *Papers for the 32nd International Conference on Salishan and Neighboring Languages*, pages 1-21.
- [Brennan, 1995] Brennan, S.E. (1995). Centering attention in discourse. *Language and Cognitive Processes*.

[Gordon et al., 1993] Gordon, P., Grosz, B., and Gilliom, L. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17(3):311-347.

[Grosz et al, 1995] Grosz, B., Joshi, A., and Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203-225.

[Turan, 1995] Turan, U.D. (1995). *Null vs. Overt Subjects in Turkish Discourse: A Centering Analysis*. PhD thesis, University of Pennsylvania.

[Walker et al., 1992] Walker, M., Iida, M., and Conte, S. (1992). Japanese discourse and the process of centering. Technical Report MS-CIS-92-32, University of Pennsylvania School of Engineering and Applied Sciences, Computer Information Science Department.

## 7. Appendix A

1 Nephi 1:10

1 Nephi 3:11

1 Nephi 3:13

1 Nephi 3:24

1 Nephi 7:1

1 Nephi 11:27

2 Nephi 11:2-3

Jacob 7:8-11

Mosiah 6:6

Mosiah 6:6

Mosiah 8:6

Mosiah 17:5-7

Mosiah 17:11-12

Mosiah 18:15

Mosiah 19:4-6

Mosiah 22:5

Alma 1:7-11

Alma 4:16

Alma 4:16-19

Alma 5:41-42

Alma 5:59

Alma 11:26-39

Alma 12:1

Alma 12:7

Alma 15:18

Alma 18:13-16

Alma 18:18-19

Alma 18:23-43

Alma 19:7-8

**8. Appendix B**

Genesis 2:19, 21-22  
 Genesis 3:1  
 Genesis 3:24  
 Genesis 4:4-5  
 Genesis 4:17  
 Genesis 4:26  
 Genesis 5:4  
 Genesis 5:22,24  
 Genesis 5:29  
 Genesis 14:14-16  
 Genesis 15:3-13  
 Genesis 18:17-19  
 Genesis 18:28-33  
 Genesis 19:27-29  
 Genesis 22:1-2  
 Genesis 27:18-36  
 Genesis 28:6  
 Genesis 29:21-25  
 Genesis 29:28-31  
 Genesis 30:28-29  
 Genesis 31:20-23

Genesis 32:24-29, 32  
 Genesis 33:1-15  
 Genesis 34:1-7  
 Genesis 35:9-14  
 Genesis 37:3-6  
 Genesis 37:13-16  
 Genesis 38:10  
 Genesis 39:2-23  
 Genesis 41:42-45  
 Genesis 45:26-27  
 Genesis 46:1-3  
 Genesis 46:29  
 Genesis 47:29-31  
 Genesis 48:17+  
 Exodus 18:14  
 Exodus 21:12  
 Exodus 21:14  
 Exodus 21:16  
 Exodus 21:18-19  
 Exodus 21:20-21

Exodus 21:26-27  
 Exodus 21:29-32  
 Exodus 33:14-15  
 Exodus 33:18-19  
 Exodus 35:30-35  
 Leviticus 8:6-12  
 Leviticus 14:19-22  
 Leviticus 13:1-14  
 Numbers 1:19  
 Numbers 5:15  
 Numbers 6:9-11  
 Numbers 7:89  
 Numbers 10:29-32  
 Numbers 12:1-2  
 Numbers 23:4-7  
 Numbers 23:11-18  
 Numbers 24:10  
 Numbers 24:20-24  
 Numbers 35:16-28  
 Deuteronomy 21:17