

# Some of the Implications of Contrastive Analysis of Morphology in English and Persian for a Translator: Based on the Theory of Information Structure

Habib Gowhary

Islamic Azad University, Ilam Branch

## INTRODUCTION

Lambrecht<sup>1</sup> following Prince<sup>2</sup> (1981) and Chafe (1976) has made a distinction between identifiable and unidentifiable referents. He explains the difference in the following terminology:

an identifiable referent is one for which a shared representation already exists in the speaker's and listener's mind at the time of the utterance, while an unidentifiable referent is one for which a representation exists only in the speaker's mind.<sup>3</sup>

The grammatical correlate of the above distinction is the formal distinction made in many languages between definite and indefinite noun phrases. In many languages, the cognitive distinction between identifiable and unidentifiable referents is morphologically expressed by definite and indefinite articles and other determiners such as *this*, *his*, and etc. It should be noted, however, this distinction between identifiable and unidentifiable referents may have different implications in different languages as far as linguistic coding is concerned. In this study, in particular it is claimed that morphological coding of mental referents affected by identifiability parameter may not be the same in English and Persian. Certain languages may have no grammatical category to code linguistically identifiability. As we will see (3) for some referents, Persian has no formal marker to show identifiability, that is why in this study a distinction is made between “(in)definite NPs/with or without marker”.

- 
1. Kund Lambrecht, *Information Structure and Sentence Form* (Cambridge: Cambridge University Press, 1994), p.77.
  2. Ellen Prince, “Topicalization and Left Dislocation: a Functional Analysis”, *Annals of the New York Academy of Sciences*, 433 (1984), pp.213-225.
  3. Kund Lambrecht, *Information Structure and Sentence Form*, pp.77-78.

According to what we have said so far regarding identifiability, it is misleadingly implied that referents are either identifiable or unidentifiable. However, Lambrecht believes that identifiability is continual in nature.<sup>4</sup> Therefore, based on the degree of identifiability, he classified the mental referents into four groups: active, semi-active, inactive, and new. The first three groups are considered to be identifiable and the fourth (new referents) is unidentifiable. Following Chafe (1987), Lambrecht emphasizes that “our minds contain very large amounts of knowledge or information, and that only a very small amount of this information can be active at any one time” (Lambrecht, 1994: 93).<sup>5</sup> Lambrecht argues that a particular concept may be in any one of “activation states”. Accordingly, Lambrecht following Chafe (1987) defines different activation states in the following terms:

“an active referents” is one that is currently lit up, a concept in a person’s focus of attention at a particular moment. “a semi-active referent” is one that is in a person’s peripheral consciousness, a concept of which a person has a background awareness, but one that is not being directly focused on. “an inactive (unused) referent” is one that is currently in a person’s long-term memory, neither focally nor peripherally active. Finally, “a new referent” is one for which the listener has no mental picture at all.”<sup>6</sup>

Based on the degree of accessibility, Lambrecht divides the semi-active referent into three minor groups which are “textually semi-active”, “inferentially semi-active” and “situationally semi-active.”<sup>7</sup> Furthermore, he made another distinction between “anchored new referents” and unanchored new referents”. Because of the reasons to be explained later, we are not dealing with these minor groups. In this study, the four main groups are dealt with. In this connection the following research questions are posed:

Question 1: Does “the degree of activation of the mental referents” affect “the morphological coding” in the two languages contrasted?

Question 2: (If the answer of the first question is YES) Is the effect of “the degree of activation of the mental referents” the same or different in the two languages involved?

---

4. Kund Lambrecht, *Information Structure and Sentence Form*.

5. Kund Lambrecht, *Information Structure and Sentence Form*, p.93.

6. Kund Lambrecht, *Information Structure and Sentence Form*, p.94.

7. Kund Lambrecht, *Information Structure and Sentence Form*, p.100.

## METHOD

### DATA

In this study, a corpus of English data together with its Persian translation was analyzed in order to investigate the effect of the degree of identifiability/activation of mental referents on linguistic coding. Another goal to be followed here is to investigate whether the identifiability parameter has the same or different effect in English and Persian. To achieve these goals, the first chapter of George Orwell's *Animal farm* together with its Persian translation by Amir Amirshahi were analyzed contrastively.<sup>8</sup> This corpus includes more than 290 clauses and 665 noun phrases in English. The Persian translation includes almost the same amount of data. The analysis of the data is based on Lambrecht's information structure theory.<sup>9</sup>

### PROCEDURES

The following steps were taken to analyze the data according to the criteria and concepts introduced by Lambrecht.<sup>10</sup>

1. The source text (English) and target text (Persian) were juxtaposed and contrasted sentence by sentence and phrase by phrase.

2. Based on the functional considerations introduced by Lambrecht, each phrase was assigned a referent.<sup>11</sup>

3. Based on functional and contextual considerations, the mental referents are classified into groups and subgroups. Following Lambrecht,<sup>12</sup> in this study the referents are divided into four main groups which are active, semi-active, inactive and new. The semi-active referents themselves fall into three subgroups which are inferential, textual and situational. Furthermore, the new referents are of two types: anchored and unanchored.

4. Each type of mental referents has a unique mode of morphological expression. In this study, a formal mode of morphological expression is identified for each type of referent. Accordingly, the following types of morphological expression were identified: definite NP, pronoun, zero pronoun, indefinite NP, and wh-words. Definite and indefinite NPs may be

---

8. George Orwell, *Animal Farm* (New York: New American Library, 1946).

9. Kund Lambrecht, *Information Structure and Sentence Form*.

10. Kund Lambrecht, *Information Structure and Sentence Form*.

11. Kund Lambrecht, *Information Structure and Sentence Form*.

12. Kund Lambrecht, *Information Structure and Sentence Form*.

used with or without formal markers such as *the, this, his* and etc. So, totally seven mode of morphological forms were identified to code the mental referents linguistically.

5. In this step, an attempt was made to make a connection between different types of referents (step 3) and their mode of morphological coding (step 4). To investigate the relationship between “the degree of the activation of referents” and “their morphological coding”,  $X^2$  test is employed to show the significance of this relation and to investigate the effect of the degree of activation on morphological coding, Z test was used to see whether the effect is the same or different in the two languages contrasted.

## RESULTS

As it was stated in (2-2), the mental referents were classified into four groups. In addition, seven modes of morphological expression were also identified. Table (1) shows the statistics of the two variables involved, that is; the degree of identifiability/ activation and different modes of morphological coding.

TABLE 1— The Frequency and Percent of the Phrases Based on "The Degree of the Activation of the Referents" and "Morphological Coding" in English

Morphological coding Degree of act	definite NP/ marker	definite NP/ no marker	definite NP/ pronoun	definite NP/ zero pronoun	definite NP/ marker	definite NP/ no marker	definite NP/ wh- words	Total
active	27	17	204	57	0	0	0	305
	8.9%	5.6%	*66.9%	18.7%	.0%	.0%	.0%	100.0%
semi-active	120	40	7	1	2	1	0	171
	*70.2%	23.4%	4.1%	.6%	1.2%	.6%	.0%	100.0%
inactive	64	30	1	0	5	5	0	105
	*61.0%	28.6%	1.0%	.0%	4.8%	4.8%	.0%	100.0%
new	5	1	1	0	60	5	11	83
	6.0%	1.2%	1.2%	.0%	*72.3%	6.0%	13.3%	100.0%
total	216	88	213	58	67	11	11	664
	32.5%	13.3%	32.1%	8.7%	10.1%	1.7%	1.7%	100.0%

$$X^2 = 9/763, P = 0/00$$

Table (2) shows the same statistics in Persian:

TABLE 2— The Frequency and Percent of the Phrases Based on "The Degree of the Activation of the Referents" and "Morphological Coding" in English

Morphological coding Degree of act	definite NP/ marker	definite NP/ no marker	definite NP/ pronoun	definite NP/ zero pronoun	definite NP/ marker	definite NP/ no marker	definite NP/ wh- words	Total
active	21	27	90	159	1	1	0	299
	7.0%	9.0%	30.1%	*53.2%	.3%	.3%	.0%	100.0%
semi-active	43	115	4	1	4	2	0	169
	25.4%	*68.0%	2.4%	.6%	2.4%	1.2%	.0%	100.0%
inactive	14	78	1	1	2	8	0	104
	13.5%	*75.0%	1.0%	1.0%	1.9%	7.7%	.0%	100.0%
new	0	4	0	0	55	13	11	83
	.0%	4.8%	.0%	.0%	*66.3%	15.7%	13.3%	100.0%
total	78	224	95	161	62	24	11	655
	11.9%	34.2%	14.5%	24.6%	9.5%	3.7%	1.7%	100.0%

$X^2 = 9/506, P = 0/00$

ACTIVE REFERENTS

As indicated in table (1), the phrases with active referents are mainly coded as pronoun in English. Out of the total number of 305 phrases which are assigned active referents, %66/9 are coded as pronoun, %18/7 as zero pronoun and %8/9 as “definite NP/ with marker.” This table shows none of the phrases with active referents are coded as indefinite NP. Table (2) shows the relationship between the degree of identifiability and different modes of morphological coding in Persian. Out of the total number of 299 phrases which are assigned active referents, %53/2 are coded by using zero pronoun. In %30/1 of these kind of phrases, pronoun is used to show the degree of identifiability. Definite NPs (with and without marker) are also used in smaller scales to express active referents. In total, %16 of active referents are linguistically represented by definite NPs. As shown in Tables (1) and (2),  $X^2$  test proved the presence of a significant relationship between the two variables ( $p = 0/00$ ). Therefore, the first question of the research is

positively proved. There is a significant relation between the degree of activation of the referents and morphological coding as far as active referents are concerned.

To answer the second question, Z test was employed to show whether “the degree of activation” has the same effect on “the morphological coding” in the two languages or not.

$$Z = \frac{P_1 - P_2}{\sqrt{P(1-P) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

In connection with the effect of the identifiability in the two languages, English pronouns used to represent active referents were compared with Persian pronouns. From the total number of 305 phrases assigned active referents in English, %66/9 are coded by pronoun and from the total number of 299 phrases assigned active referent in Persian, %30/1 are coded by pronouns. The calculated Z ( $Z = 9/1$ ) shows that the obtained Z is more than  $1/96$ , so the attested difference between English and Persian is statistically significant ( $P < \%05$ ). In other words, English and Persian speakers differ in their use of “pronouns” to express the active referents.

#### SEMI-ACTIVE REFERENTS

Table (1) indicates out of the total number of 171 phrases which are identified to have semi-active referents, %70/2 are expressed by “definite NP/ with marker” in English. These referents are rarely expressed by pronouns and indefinite NPs. On the other hand, table (2) shows that out of the total number of 169 phrases which are thought to have semi-active referents in Persian, %68 are expressed by “definite NP/ no marker”, %25/4 by “definite NP/ with marker and about %3/6 by indefinite NP. Totally, %93/4 of such referents are represented by “definite NP” in narrative texts.  $X^2$  test shows that the relation to be significant in both languages ( $P = 0/00$ ).

As for semi-active referents,  $P_1 = \%70/2$  and  $P_2 = 25/4$ . The statistics of the calculated Z test showed that the difference between English and Persian in connection with the use of “definite NP/ marker” is statistically significant ( $P < \%05$ ). Accordingly, to express semi-active referents by using “definite NPs/ with marker” in narrative texts, English speakers are different from Persian speakers. This difference is statistically significant. Again, to answer the second question: English mode of morphological coding is different from Persian as far as semi-active referents are concerned.

#### INACTIVE REFERENTS

As shown in table (1), from the total number of 105 phrases which are assigned inactive (unused) referents in English, %61 are coded by using “definite NP/with

marker”, %28/6 by “definite NP/ without marker”, and %9/6 by indefinite NP. In total, %89/6 are represented by definite NPs. Table (2) shows that in Persian, from the total number of 104 phrases which are identified to have inactive (unused) referents, %75 are expressed by using “definite NP/ without marker”, %13/5 by “definite NP/ with marker”, and %9/6 by indefinite NP. In total, %88/5 are represented by definite NPs. To address the first question,  $X^2$  test shows that the relation as to be significant in both languages ( $P=0/00$ ).

In connection with the second question for inactive referents,  $P_1= %61$  and  $P_2= %13/5$ . The calculated Z test showed that the difference between English and Persian in connection with the use of “definite NP/ marker” to express inactive referents is statistically significant ( $P < %05$ ). Accordingly, to express inactive referents by using “definite NPs/ with marker” in narrative texts, English speakers are different from Persian speakers. To answer the second question: the effect of the degree of activation on morphological coding has different repercussions in English and Persian as far as inactive referents are concerned.

#### NEW REFERENTS

Table (1) indicates that out of the total number of 83 phrases with new referents in English, %72/3 are coded morphologically by “indefinite NP/ with marker”, %13/3 by wh-words, %6 by “indefinite NP/ without marker”. “definite NP” is only used in %6 to express new referents. Totally, %91/6 of the new referents in English are expressed by “indefinite NP”. On the other hand, table (2) shows that out of the total number of 83 phrases assigned new referents in Persian, %66/3 are represented in the language by “indefinite NP/ with marker”, %13/3 by wh-words, %15/7 by “indefinite NP/ without marker”. “definite NP” is rarely used to express new referents in Persian. Only, in %4/8 of the new referents, “definite NP” was employed to express them. Totally, %95/2 of the new referents in Persian are expressed by “indefinite NP”. As the above referents,  $X^2$  test proves the relation between the two variables as to be significant.

To address the second question in connection with the new referents, English “indefinite NPs/ with marker” used to represent new referents were compared with Persian “indefinite NPs/ with marker”. From the total number of 83 phrases assigned new referents both in English and Persian, %72/3 are coded by “indefinite NPs/ with marker”. In Persian this percent is %66/3. The calculated  $Z^* = %74$ . The obtained  $Z^*$  is lower than  $1/96$ , so the attested difference between English and Persian is *not* statistically significant

( $P > \%05$ ). To answer the second question: the identifiability effect has led English and Persian speakers to use “indefinite NPs/ with marker” almost in the same way to express new referents in narrative texts.

## DISCUSSION

According to the descriptive statistics presented above, pronoun and zero pronouns respectively in English and Persian are preferred morphological modes of expression to represent active referents. Another finding of the analysis of the data is the fact that morphological coding is affected by the degree of activation, in particular; active referents, in this case, are coded by using (zero) pronoun in both languages.  $X^2$  test proved the relation to be significant. In addition, the results show that the effect of the degree of activation of the referents has led the English and Persian speakers to use different modes of morphological expression; as stated before, pronoun in English and zero pronouns in Persian were identified to express active referents. The reasons for this difference in English and Persian are beyond the scope of this study, so this is not dealt with here.

To express semi-active referents in narrative texts, English speakers tend to use “definite NPs/ with marker”. In other words, “definite NPs/ with marker” are preferred morphological forms to express such referents. Based on the patterns emerged in the data both in English and Persian, rarely are semi-active referents coded by pronoun or indefinite NP. Any exception must have some pragmatic and functional reasons. On the other hand, definite NPs are preferred morphological forms to represent semi-active referents in Persian translation. However, there is a minor difference in this connection. As it was stated, in English definite NPs are mainly accompanied with formal morphological markers such as *the*, *this*, *his* and etc, but in Persian, definite NPs are usually coded without such formal markers.

Almost the same pattern that emerged in the morphological coding of semi-active referents in both languages is also seen in the morphological coding of inactive referents. In both languages, “definite NP” is the morphological preferred form to express inactive referents. Like semi-active referents, there is a difference in this connection between the two languages, that is; while in English definite NPs to represent inactive referents are accompanied with formal markers of definiteness such as *the*, *this*, *his* and etc., in Persian the use of such markers with NPs is not the dominant tendency. Accordingly, it can be



claimed that there is a difference in the morphological expression of inactive referents between English and Persian; formal markers of definiteness to express semi-active referents are more frequent in English than Persian. However, it cannot be denied that in both languages, inactive referents are coded morphologically by using “definite NP”.

In the discussion of active referents, it was concluded that these referents are mainly coded by weak morphological forms such as pronouns. Now, it is suggested that new referents are coded by strong morphological forms. Indefinite NPs which are the preferred forms to express new referents are morphologically considered to be strong based on iconicity principle. In contrast to pronouns, NPs contain more linguistic material, hence strong.

It was concluded above that there was a strong correlation between active referents and the use of morphological weak forms such as (zero) pronouns on the one hand and between new referents and the use of morphological forms such as indefinite NPs on the other hand. Now, a point worth-mentioning here is related to morphological coding of semi-active and inactive referents. Table (1) shows that in total, %93/6 of the semi-active referents and %89/6 of inactive referents in English are coded by “definite NP/ with or without marker”. Table (2) shows almost the same pattern in Persian. Therefore, it is proved both in English and Persian “definite NP” is the preferred morphological form. However, a question which is logically raised here is “why are both semi-active and inactive referents expressed by the same morphological form?” This question is a challenge for a theory which claims that every kind of referent with a certain degree of identifiability has a unique mode of expression. On the other hand, this apparent problem challenges the authenticity of the classification proposed by Lambrecht,<sup>13</sup> Chafe (1976), and Prince<sup>14</sup> (1981). If this classification of the mental and cognitive referents is psychologically real, then it must be reflected in language. However, the researcher has a solution for this problem. It must be noted that in this study, we have investigated the effect of the degree of idetifiability of the referents on morphological coding. In order to solve the above problem, it is needed to investigate the effect of the degree of identifiability on other areas of language such as phonology and syntax. Accordingly, it must be emphasized that a native speaker has various tools such as morphology, syntax and phonology to express his mental assumptions. In the above problem, it was stated that English and Persian speakers used the same tool to express semi-active and inactive

---

13. Kund Lambrecht, *Information Structure and Sentence Form*.

14. Ellen Prince, "Topicalization and Left Dislocation: a Functional Analysis."

referents. However, it must be noted that this ambiguity at morphological level is eliminated by different patterns of accentuation at phonological level. It is true that both semi-active and inactive referents are coded morphologically by the same mode of expression, but careful analysis of the data shows that the “definite NPs” which are used to code inactive referents have higher chance of receiving accent in contrast with “definite NPs” which are used to express semi-active referents. Table (3) shows the patterns of accent assignment in English and Persian contrastively.

As shown in the table, the phrases with more unidentifiable referents have the higher chance of receiving accent. Accordingly, a larger number of phrases with semi-active referents are accented than phrases with active referents and in the same vein a larger number of phrases with inactive referents are accented than phrases with semi-active referents. Therefore, it can be concluded that there is an interaction between different components of language to code mental referents. Based on what we have said so far, the researcher suggests the following principle

TABLE 3— The Frequency and Percent of the Phrases Based on “Accent Status” and “The Degree of the Activation” in English and Persian

Degree of activation	English		Total	Persian		Total
	Accent status			Accent status		
	No	Yes		No	Yes	
Active	268	37	305	263	35	298
Percent	*87.9%	12.1%	100.0%	88.3%	11.7%	100.0%
Semi-active	78	93	171	75	94	169
Percent	45.6%	*54.4%	100.0%	44.4%	*55.6%	100.0%
Inactive	34	71	105	32	71	103
Percent	32.4%	*67.6%	100.0%	31.1%	*68.9%	100.0%
New	7	76	83	6	76	82
Percent	8.4%	*91.6%	100.0%	7.3%	*92.7%	100.0%
Total	387	277	664	376	276	652
Percent	58.3%	41.7%	100.0%	57.7%	42.3%	100.0%

English  $X^2 = 23/79$ ,  $df = 3$ ,  $P = 0/00$

Persian  $X^2 = 24/14$ ,  $df = 3$ ,  $P = 0/00$

### The Principle of Interaction between Components of Language (PICL)

“If there is an ambiguity regarding the linguistic coding of mental referents in one component of language, we can eliminate this problem by referring to other areas of language.”

### CONCLUSION

Based on the above discussion, it can be claimed that the morphological coding in both English and Persian is affected by the degree of identifiability of the mental referents; in particular, identifiable referents were found to be represented by weak forms and unidentifiable referents by strong forms. In addition, the effect of identifiability of the referents on morphological coding has not been always the same in the languages contrasted, e. g. phrases with active referents are coded morphologically by pronoun in English and zero pronoun in Persian or phrases which are assigned semi-active referents are expressed mainly by “definite NP/ with marker” in English and “definite NP/ without marker” in Persian. What these findings imply for a translator is important: literal translation of morphological elements of SL has a less chance of success. Second, it seems a formal training of information structure is becoming more urgent for a translator.

Finally, PICL implied that to code the mental states of the referents completely, all the components of language are involved simultaneously.

### REFERENCES

- Prince, Ellen (1984), “Topicalization and Left Dislocation: a Functional Analysis.” *Annals of the New York Academy of Sciences*, 433 (1984), pp.213-225.
- Orwell, George. (1946), *Animal Farm* (New York: New American Library), [translated by Amir Amirahsahi], *Translation of Animal Farm* (Tehran: Amir-e-kabir Publications, 2009).
- Lambrecht, Kund. (1994), *Information Structure and Sentence Form* (Cambridge: Cambridge University Press).

