

# A Contextual Approach to the Study of Discourse Anaphoric Expressions: On the Probabilistic Nature of Theoretical Predictions

Mozhgan Madhkhan<sup>1</sup> & Mojtaba Mousavi<sup>2</sup>

<sup>1</sup> Faculty of Foreign Languages, Yazd University, Yazd, Iran

<sup>2</sup> Faculty of Medical Science, Isfahan University, Isfahan, Iran

Correspondence: Mozhgan Madhkhan, Faculty of Foreign Languages, Yazd University, Yazd, Iran. E-mail: m.madhkhan7@gmail.com

Received: June 15, 2017 Accepted: September 8, 2017 Online Published: September 10, 2017

doi: 10.5539/elt.v10n10p77 URL: <http://doi.org/10.5539/elt.v10n10p77>

## Abstract

Anaphoric expressions are among the most frequent language forms which depend on context for their resolution. Among efforts made in theorizing referential choice, distance approaches take into account how accessibility/continuity is reflected by choice of referring expressions. The thing is that individual's choices are made under major and minor influences which affect the variable predictability of anaphoric expressions. In order to investigate the variability of individual writings in a free writing task (the present paper aims to investigate the variability of individual writings in terms of the use of anaphoric expressions in a free writing task). Narrative data was collected from 10 native writers (of English/Persian in University of Isfahan) by presenting them a seven-minute soundless action movie and asking them to narrate its story. The results show a fair degree of diversity in varying frequencies of pronouns and Full NPs. It is suggested that diversity might be caused by two factors: referential strategies of text producers who consider the helping role of context as well as the degree of referential distance, referential ambiguity and salience.

**Keywords:** discourse anaphoric choice, referential idiosyncrasy, contextual approach, referential choice

## 1. Introduction

Anaphoric expressions are defined as linguistic items which act as formal alternatives to full forms (Allan, 1986, 1995). Anaphora refers to either the linguistic referential device used, the extra-linguistic entity referred to or the relation that can be assumed between co-referring pronouns and their antecedents. They are among the most frequent language forms in discourse (Almor, 2000). It is assumed that anaphoric expressions are not only means to guide discourse reference but also the concatenation of referential items is patterned to display the cognitive tendencies of the individual writer to convey the personal and general pragmatic intents and the structure of the discourse. Thus, a pronominal expression may convey information about what point of view is to be taken on the world as a discourse expresses (Clancy, 1980). It maybe that using a pronoun to refer to an entity that is not already salient makes that entity more salient than it would be used as a full noun phrase (NP). According to Gernsbacher (1989), reduced anaphoric devices enhance the activation of their antecedents; that is why concepts to which the reduced devices refer are more accessible and thus identified more rapidly. In order to account for how accessibility is marked by referring expressions, Distance Approach is suitable to be taken.

Those linguistic attempts which are based on Distance Approach are categorized into 3 groups: Referential Choice, Topic Continuity and Accessibility Theory (Clancy, 1980; Givón et al., 1983; Ariel, 1988, 1990, 1991). The functional studies dealing with cognitive effects of distance on accessibility of referent deal with accessibility at discourse level. By application of a given/new strategy to discourse pronouns, Clark and Sengul (1979) suggested that the more relevant and recent the referent, the less time is necessary to access the memory and link new information to it.

Discourse reference is seen as an achievement of the process of Referential Choice: each time the speaker should mention a referent, he/she has a number of options at his/her disposal, such as demonstratives, third person pronouns, full NPs, etc. The speaker chooses one of them according to certain rules which are a part of the system of language production. Production-oriented accounts of reference are much rarer in literature than comprehension-oriented ones. For some examples, see Dale, (1992); Strube and Wolters, (2000). Clancy (1980)

conducted a detailed investigation of Referential Choices made by twenty English and twenty Japanese speakers recalling the story after seeing the pear film (Chafe, 1980) to examine linguistic and extra-linguistic sources for referent retrievability. Clancy's analysis of pear stories was one of the first comprehensive attempts to verify the function of referential distance and potential ambiguity as the universal cognitive constraints on determining the arrangement of anaphoric expressions. The constraints are originally proposed by psychological studies as parameters along which the capacity of short-term memory can be measured. The amount of elapsed time between two mentions of a particular and potential ambiguity are assumed to be correlated with accessibility measured by means of clausal and sentential distance representing a rough general indication of the amount of cognitive activity. Referential salience is not measured, but the examination of several narrative pieces revealed that referential salience attracts focus of attention, and optionally motivates unusual Referential Choices. However, the way the effect of ambiguity is screened out in measuring the referential distance of the nominal distributions in the measurements is not clarified.

Clancy (1980) concluded that Referential Choice appears to be the result of an integrated effect of several factors, the interaction of which impacts on the diversity observed among individual performances. Discourse structure, cognitive constraints, optionally marked referential salience, point of view, world shifts and individual strategies are a variety of factors each sharing a role. Therefore, it is by no means easy to decipher the implication of Referential Choice. A variety of factors may interfere in introducing and maintaining reference to a character in an oral performance.

Several studies have appeared proposing that Referential Choice is directly pertinent to the more general cognitive area of working memory and the working memory process of activation (Tomlin & Pu, 199; Chafe, 1994; Givón, 1995; Kibrik, 1991, 1996, 1999; Cornish, 1999). The assertion that Referential Choice is controlled by memory processes is in line with psycholinguistic frameworks of such researchers as Vonk et al. (1992), Clifton and Ferreira (1987), Gernsbacher (1990), with the cognitively oriented approaches of the Topic Continuity research (Givón, 1983), Centering theory (Gordon et al., 1993), Accessibility theory (Ariel, 1990), Cognitive grammar (van Hoek, 1997), and Givenness hierarchy (Gundel et al., 1993), along with some computational models employed by (Botley & McEnery, 2000).

The notion of "Referential Choice" puts an emphasis on the procedural, dynamic nature of reference in discourse. Moreover, it is clearly production-oriented: Referential Choice is the process carried out by the writer/speaker. For each act of Referential Choice, the speaker selects a formal device to codify the referent he has in mind. In contrast, "anaphora" is normally perceived as a more static textual case, as a relationship between two or more pieces of the text (Grüning & Kibrik, 2002).

Among major linguistic attempts focusing on referential distance made after Clancy, Topic Continuity (TC) (Givón et al., 1983; Givón 1989, 1990) and Accessibility Theory (AT) (Ariel 1985, 1988, 1990, 1991) are regarded as two major discourse functional cognitively-based studies. The contribution of TC studies is primarily attributed to their attempts to classify grammatical devices (code quantity), syntactic case roles (linear order) and animacy of referents of linguistic elements within clauses in relation to TC according to the text-based observations of discourse topical referents. Givón (1983, 1989, 1990) introduces TC, known also as the traditional distance theory, dealing with the discourse factors that determine the choice of referring expressions in a macro-analysis of the topic. The discourse topic is the most recurring participant of discourse, or in other words, the main protagonist of the narrative, and is marked according to the Topic Continuity principle leading to referential predictability. The major assumption of TC is that code quantity (phonological length) and linear order (syntactic case roles) are sensitive to the degree of predictability which is itself determined by referential distance, ambiguity and persistence. Two important ingredients of Topic Continuity, namely code quantity and linear order, are susceptible to the major features of topicality, namely continuity and salience. The thrust of TC has been to demonstrate that the atomic view of topic and the functional ordering of topic and comment on a sentence level should be replaced with a new and more encompassing discourse-oriented wisdom. Discourse generally deals with characters who perform actions, have internal states and are acted upon by forces in their environment. At any point in a discourse, one character is usually more important than others. Pronouns tend to be used when their discourse antecedents are highly available in memory and are important. The preferred referent for a pronoun might be the entity whose representation in discourse model is the most salient at the time the pronoun is encountered.

Accessibility Theory (AT), on the other hand, is an attempt to bring context studies, psycholinguistic findings of the processing of anaphoric expressions and linguistic presuppositional contributions into one common model. The model is based on the premise that accessibility markers have contextual correlates and that this correlation is by no means arbitrary.

AT (Ariel, 1985, 1988, 1990, 1991) is the approach to the effect of distance on accessibility of entities in discourse originating from the previous research on reference in terms of presuppositionality, context type theories and psycholinguistic studies dealing with processing procedures for referring expressions (hereafter RE's). Salient referents gain growing importance and their subsequent mentions must be viewed as anaphoric. The basic assumption is that mental representations are not equally accessible at various stages of the discourse and these representations are coded in language properly. That is why accessibility is a more comprehensive term than givenness and is a better replacement for presuppositionality. AT deals with accessibility signaled by the RE's actually used rather than context; not only the discourse optional distributional possibilities regarding the use of RE's are claimed to be accountable by AT, but also the formal grammatical options languages are allowed to use in referring to entities can be addressed by it. Such a model claims that the use of any particular expression, far from being a random choice, is closely connected to the level of accessibility or activation that the mental representation of the referent is assumed to have in the addressee's mental model of the discourse under construction. On the basis of the various cognitive status which are assigned to discourse referents, these models provide a formal basis that makes it possible to justify and even to predict the presence or use of a particular marker (Fossard & Rigalleau, 2002). Henceforth, when the referent targeted is assumed to be highly accessible, highly active in the addressee's mental discourse representation, a morphologically and phonologically attenuated form such as a zero and unaccented third-person pronoun should be used, a form whose use signals precisely the 'in focus' status of the referent (in Gundel et al.'s terminology). On the other hand, where the referent is not in focus, but is 'activated' (in Gundel et al.'s terminology) or enjoys a level of 'medium accessibility' (in Ariel's terminology), a demonstrative pronoun or accented third-person pronoun, whose use codes this cognitive status, tends to be used. Finally, lexically and accentually more substantial expressions (such as definite full NPs or proper nouns) are expected to indicate referents bearing a low level of accessibility.

The functional approach contends that the choice of a referring expression is determined by a number of factors and is a reflection of the accessibility of referents. Two major factors influencing Referential Choice are salience within discourse boundaries and referential distance. Psycholinguistic research has dealt with the effect of distance, referential salience and organizational structure on Referential Choice.

Choosing anaphoric expressions is not a random process but a patterned process. It does not happen at sentential level but at textual and topical levels. A discourse is ideally about some central concepts or topics serving as the central index point for referring. Upon designation of topic, sentences are believed to be about topic and their interpretation of references are constrained to topic until it shifts. The natural antecedent for the pronoun 'his' in example 1 would be *Mike* if sentence 4 were isolated. However, in the context of discourse, 'his' is more likely to refer to *Bruce*, the text's main protagonist. Hence, it could be understood that in the context of discourse, the interpretation of the pronoun is indeed constrained by the discourse topical structure. This is evidence for the claim that the production and interpretation of pronominal expressions must be studied in the context of discourse.

Example 1:

- 1- I want to have a meeting this week.
- 2- Bruce will be the guest lecturer.
- 3- He will speak on slavery in ant colonies.
- 4- Mike wants to read *his* report before the talk.

Texts are believed to be stored in episodic memory as files hierarchy constructed. Each file has a label and the labels are the means by which files are activated. File labels are topics by which texts are retrieved, stored and processed (Given, 1990). File labels are mental representations of a coherent text which is constructed hierarchically with the lowest node level being the thematic paragraph. In this sense, salient characters receive greater focus while interacting with other referents. In the following example, 'Mary' is not pronominalized because of the primary salience and hence focus is given to 'Jim'.

Example 2:

*Jim* insisted on helping *Mary* with the typing. Although it was a simple task, *he* managed to make it seem like a major ordeal. By the time *he* finished three pages, *Mary* had finished the other ten and was ready to leave.

The features characterizing topic are those of; being indefinite, given, salient and the most often pronominalized entity, the subject of most clauses so statistically posited to the fore, dynamically developed, preferably the human referent and the most frequently mentioned entity of discourse. It's the topic that can determine what

constitutes the unit of text for the boundary of discourse which is marked by the identification of a topic shift and consequently action shift (Hartig, 1977; Longacre, 1979; Gioven, 1983).

Referential topicality is reminiscent of the concept of “empathy” (Kuno & Kaburaki, 1977). Empathy indicates that the speaker, in describing a particular event or state, may make certain lexical and syntactic choices indicating that he is identifying with one of the characters being referred to. In example 3, the speaker has an objective point of view in (a), while he takes a standpoint of *John* in (b) and takes *Mary’s* standpoint in (c). The character being sided for is pronominalized while characters receiving empathy are mentioned explicitly. The influence of character salience and empathy is realized in effect as a general tendency.

Example 3:

- a) *John* hit *Mary*.
- b) *John* hit his wife.
- c) *Mary’s* husband hit her.

Choice, acquisition and comprehension of pronominal and zero anaphora are crucially influenced by the perceived distance between pronouns and co-referential antecedents. Sengol’s finding (1979) is consistent with the spirit of distance approach. Pronouns are likely to be more difficult to understand when the representation of their referents might be supposed to be distant from them.

The fact that pronouns usually refer to the most accessible elements in discourse is presupposed by such computational approaches for resolving intersentential pronominal anaphora. In these approaches, the degree of accessibility of elements in the addressee’s discourse model is connected to their degree of salience, so that the most salient discourse elements are also the most easily accessible ones (Navarretta, 2002).

## 2. Idiosyncrasy in the Arrangement of Anaphoric Expressions

Cognitive constraints affect the predictability of anaphoric expressions in discourse. These expressions are determined not only with recency and discourse structural factors, but also by the writer’s stylistic freedom. Two distinct functions of a full NP include: (a) maintaining clarity in the presence of referential ambiguity and (b) conveying optional pragmatic intents of the text producer. Taking into account the unpredictability caused by stylistic freedom of the individual producers, an analysis based upon the examination of the use of referring expressions in discourse can only lead to probabilistic results. The study of context in relation to co-reference resolution leads to an improved post hoc explanation of how referents are identified.

The role of any referring expression is to enable the addressee to pick out the corresponding item from a reference set as economically as possible; so, a correlation is assumed to exist between the use of a given referring expression and the predictability of the intended antecedent, determined by the cognitive effect of either referential distance or focus within the episode boundaries. Enkvist (1978) stated that “the manner in which the text has been put together out of such [episode] units can contain stylistic characteristics.”

The identity of a given referent is connected to both informativity of context and the informativity of the referring expression used. When a given referent is not easy to identify through context, a full NP is used to avoid referential ambiguity; and when context reveals the identity of the referent in question, referential distance does not constrain the use of pronoun.

Accessibility is meaningful once a discourse view takes into account the possible pragmatic inferences within a world evoked by the discourse. A context-based accessibility approach is far-reaching and assumes a stronger function for context than has hitherto been assumed.

## 3. Methodology

In order to investigate the variability of individual writings in a free writing task, narrative data was collected from ten speakers (male and female) by presenting them a seven-minute soundless action movie and asking them to narrate its story. Age, social class and education level were assumed to be roughly constant, because subjects were attending the same class at University of Isfahan. Since the time allocated for writing the stories was limited, the subjects made their utmost effort to write as much as they could. Narratives written by individual writers contain stylistic diversity and are of variable lengths.

The following is one subject’s version of the story:

Steed was driving along a road when he conies across another car parked across the road. Seeing no driver, lie looks around and sees a quarry nearby. At the bottom of a deep pit, he saw what appeared to be a body lying at the bottom. He climbed down a ladder placed handily and while he was checking the body, the ladder was

removed. He tried several times to scale the side of the quarry but failed. An earth moving vehicle started up on the top of him and someone proceeded to push shovels of earth into the quarry immediately above him. After three loads of earth had been dumped, the driver, wearing red socks and black shoes and a blue suit, comes to the edge. Amongst the dumped earth he sees Steed's bowler hat so he descends into the quarry. Meanwhile, Steed, still neatly dressed in his black suit etc., had been hiding and attacks the bulldozer driver from behind. They fight. Meanwhile the bulldozer begins to move forward on its own and eventually tips over into the quarry. The bulldozer driver was losing the combat. Steed climbs up over the bulldozer to the top of the quarry and picks up his bowler, pushes out the dirt and dusts it down and puts it on his head.

In the above narrative, *the subject of driving, coming across, seeing, looking in a nearby quarry, seeing a body, mid climbing down*, is Steed without any other referent competing. Then the action of *removing the ladder* functions to introduce a new character into the action part indirectly. Steed's resumption of actions begins by trying to *scale the side of the quarry* and failing. Again a vehicle represents a mysteriously represented character; *"a tractor starts up on top of him"*. Now the second character is introduced by *someone*; and he is represented by further actions of *proceeding to push shovels of earth*. Steed is still under focus and is coded by him in the object position. The indefinite pronominal expression, *someone*, is changed to *the driver wearing red and black shoes and a blue suit*, and is represented by *'coming above the edge'*, he is now in focus and Steed appears to receive secondary attention at this stage. The expressions *seeing Steed's bowler hat and descending* predicate the sinister man. A full NP, Steed, is used to make a switch subject. The referent is represented by actions such as hiding nearby and attacks; the latter is associated also with the driver who is the recipient of the action of attack. The different types of full NP expressions may be for expressing an attitude: the driver of the bulldozer is marked as the bad man, the sinister man and so forth. This diversity tends to be for stylistic felicity. The verb fight predicates both of the men, referred to by the plural pronoun they. Focus is then shifted to the bulldozer which is represented by beginning to move forward and tipping over into the quarry. The action shifts again, so the bulldozer driver takes subject position being represented by losing the combat. The word combat indicates the presence of Steed, although he is eliminated from mentioning in the sentence; this is the property of the type of predicate that makes it possible to refer to the driver in the subject position as the recipient of an action the agent of which is not mentioned. In the end, the conquest is achieved and the conclusion part begins. A switch again is made to Steed in the conclusion part; he becomes the subject of climbing up, picking up the hat, pushing out the dirt, lusting down, and putting the hat on his head.

The referent who is introduced with an indefinite NP seems not to persist. Focus is allocated with more difficulty for referents who are introduced initially by an indefinite NP, when there is competition and the context contains interactive events. In the above story, the antagonist of the story is first introduced indirectly, then an indefinite pronominal expression is used and is again reintroduced by several different terms: first he is introduced indirectly through a passive construction, the ladder is removed, then the clause an earth moving vehicle starts up indirectly implies his action so again the presence of the driver is pragmatically inferred; the next explicit mention is made by someone, the fourth mention is made by a much longer expression: the driver wearing red socks and black shoes and a blue suit, the fifth and the sixth mentions are made by tractor driver. Contrast this character's introduction and reinstatement with the main character, Steed, whose introduction and reinstatement is done through the use of a simple name and for whose further mentions either the same full NP type is used or a pronoun.

As an effective stylistic strategy, it is possible to leave the antagonist unmentioned through the use of passive constructions and the use of a bridging inference, e.g. using tractor instead of the man driving it, the strategy provides significant help in the task of referring to the characters in potentially ambiguous environments. One useful stylistic technique in elliptical marking of a referent is to leave an entity unmentioned (the strategy tends to be used as a device to create suspense for the reader) while his presence is felt. In the following examples, the referent who is the center of focus is not mentioned, but one can observe the implicit presence. So focus may not necessarily be made explicit by a pronominal mention; there may not even be an ellipted NP slot for some mentions. A strong expectation is created for the reader that someone who is not mentioned on the surface is present in the discourse at this stage of the story. Formal treatments ignore some of these cases of elliptical co-reference; only pragmatic inferences within the story world handle the resolution in (b) & (c).

- 1) a. The ladder is being slowly pulled up.
- b. A bulldozer starts coming to the edge of the quarry.
- c. The ladder has gone.

While there are clear examples of long distance co-reference by a pronoun dominated by focus of attention, it

can hardly be claimed that all instances of anaphor resolution can be accounted for in terms of the cognitive effect of focus within episode boundaries. The main issue in accounting for focus of attention is that focus is optionally marked by the use of anaphoric expressions. Example two illustrates the use of a long distance pronoun to resume reference to an entity, a body, after a long gap:

2). He investigates at a nearby quarry, and notices what looks like a body in the quarry [6 clauses intervening ...] He finds that it is merely a coat and hat in the quarry.

It is interesting to note that whereas a body in the example above is pronominalizable after six clauses intervening; in spite of its immediate previous mentions, nowhere in the data collected is quarry pronominalized. While both, the body and quarry are identifiable through the informativity of context in most of the NP slots they occur in, one is pronominalized but the other is not. How can we account for this difference? One answer is that inanimate entities used as locatives may be identifiable through context; however, there is stylistically very little preference for using pronouns in marking them. The claim is that a referent may be identifiable through predications made of it, but the type of referring expression used for marking its identifiability may appear unpredictable and stylistically variable. The difference should be taken into account in dealing with how referring expressions are arranged in discourse.

In the following example the man becomes the body in the next clause. This kind of alternation is possible when the entire story is viewed and when a shared knowledge of the world spoken of enables the bridging between the two semantically different terms which are co-referential. Through a pragmatic inference, it is possible to think of the man as a body in the next stage of the story. The inferences are based on who is likely to be associated with which action within the limits of a world spoken of.

3). A bolting man gets out of the bulldozer {O} believing that the man has been crushed, a hat lies on the ground where the body apparently is buried.

There is persuasive evidence to claim that it is the context, rather than the pronominal expression, that reveals the identity of some referents. In seven, it is not the pronoun that reveals the identity of the referent in the ambiguous environment. The reference by the pronoun is not resolved until the entire clause is uncovered:

4). He, [the blue suited man] believed he, had been successful and descended into the pit.

The anaphoric relation between the pronoun he in the second clause of the sentence and its previous antecedent marked as He, is resolved by the informativity of the context within which the two mentions are located; in a narrow sense the complementary expression successful is associated with the referent of He, as we might get a different impression if the clause read:

5). He, (the blue suited man) believed he<sub>2</sub> had been buried under the soil and he descended into the pit.

The examples below further show the importance of the clausal predicates in the identification of the referents marked by pronominal devices. In 6, he is used for marking Steed, then the same referent is marked by Steed, rather than he, and finally in the last clause, context determines the referent of the slot; so the choice of a marking device is more optional. In 7, the use of ellipsis removes the effect of potential ambiguity.

6). He (Steed) evaded the dirt and when the driver of the dozer stopped to see whether Steed was buried, [zero]/he/Steed jumped upon him from behind.

7). As he (the evil man) leaves to pick up what he thinks is the hat visible indication of Steed, the latter jumps on top of him, [zero] engaging him in yet another titanic confrontation.

#### 4. Result

As the results of a simple statistical comparison show, a wide diversity is observed which is connected to several factors including the possible effect of stylistic freedom. Even different frequencies of subject switch and varying selection of types of verb phrases may bear different results regarding the frequency of pronouns and full NPs used. Table one represents the consecutive mentions of the quarry where most of the actions in the story take place. The reason for the explicit reference to this referent cannot be that the location of the quarry is unidentifiable: the plan of the story and the explicit location of all the actions highlight it; but despite lack of ambiguity, it appears stylistically not desirable for this entity to be pronominalized. The other feature of the narratives which makes them less homogenous is the great difference across subjects in the frequency of mentions of the entities (3 to 13 times in the case of the quarry). The variability in the number of entities introduced and the frequency of mentioning them both impact on the frequency of types of anaphoric expressions used.

Table 1. The full NP types used to refer to the protagonist of the story, Steed/a man as contrasted with the antagonist, a man/the driver of the tractor. Pronominal mentions are extracted from comparison

Columns 1 & 2 display how the two characters of the story are referred to by the subjects 2,4,6,8 and 10		Columns 3 & 4 display how the two characters of the story are referred to by the subjects 3,5,7,9 and 11.	
<u>SUBJ 1</u>	<u>SUBJ 1</u>	<u>SUBJ 6</u>	<u>SUBJ 6</u>
<ul style="list-style-type: none"> <li>• Roy</li> <li>• Roy</li> <li>• Roy</li> <li>• The dead man</li> <li>• Roy</li> </ul>	<ul style="list-style-type: none"> <li>• The noises of a tractor</li> <li>• The tractor</li> <li>• The driver of the tractor</li> <li>• The driver</li> </ul>	<ul style="list-style-type: none"> <li>• The black suited man</li> <li>• The black suited man</li> <li>• The black suited man's</li> <li>• The black suited man</li> <li>• The black suited man</li> </ul>	<ul style="list-style-type: none"> <li>• The blue suited man in a tractor</li> <li>• The blue suited man</li> </ul>
<u>SUBJ 2</u>	<u>SUBJ 2</u>	<u>SUBJ 7</u>	<u>SUBJ 7</u>
<ul style="list-style-type: none"> <li>• J Steed</li> <li>• Steed</li> <li>• Steed</li> <li>• Steed</li> </ul>	<ul style="list-style-type: none"> <li>• Someone</li> <li>• The driver of tractor</li> <li>• His opponent</li> </ul>	<ul style="list-style-type: none"> <li>• A man dressed in a suit</li> <li>• The man</li> <li>• The first man who is not really buried under the soil</li> </ul>	<ul style="list-style-type: none"> <li>• A bulldozer (+ driver)</li> <li>• The man driving the tractor also dressed in a suit wearing dark glasses and looking sinister</li> <li>• One man (the writer is dubious)</li> <li>• The tractor driver</li> </ul>
<u>SUBJ 3</u>	<u>SUBJ 3</u>	<u>SUBJ 8</u>	<u>SUBJ 8</u>
<ul style="list-style-type: none"> <li>• Steed</li> <li>• Steed</li> <li>• Steed</li> <li>• Steed's bowler</li> <li>• Steed</li> <li>• Steed</li> </ul>	<ul style="list-style-type: none"> <li>• A bulldozer is being used</li> <li>• A man in a blue suit with red socks</li> <li>• The other man</li> </ul>	<ul style="list-style-type: none"> <li>• A man</li> <li>• The man</li> <li>• The body</li> <li>• The other man</li> </ul>	<ul style="list-style-type: none"> <li>• A bulldozer (driver)</li> <li>• The driver of the tractor</li> <li>• The bolting man</li> <li>• The bolting man</li> </ul>
<u>SUBJ 4</u>	<u>SUBJ 4</u>	<u>SUBJ 9</u>	<u>SUBJ 9</u>
<ul style="list-style-type: none"> <li>• Steed</li> <li>• Steed's hat</li> <li>• Steed</li> <li>• Steed</li> </ul>	<ul style="list-style-type: none"> <li>• Someone</li> <li>• The driver wearing red socks and black shoes and a blue suit</li> <li>• The tractor driver</li> </ul>	<ul style="list-style-type: none"> <li>• A man</li> <li>• The man</li> <li>• The man</li> <li>• The man</li> <li>• The man</li> <li>• The man's hat</li> <li>• The good guy</li> <li>• The good guy</li> <li>• The good man</li> </ul>	<ul style="list-style-type: none"> <li>• An engine</li> <li>• This engine</li> <li>• The front of the tractor (+ drive)</li> <li>• The tractor (+ driver)</li> <li>• The person driving the tractor with gloved hands and maroon socks</li> <li>• The bad guy</li> <li>• The bad guy</li> <li>• The bad guy</li> </ul>
<u>SUBJ 5</u>	<u>SUBJ 5</u>	<u>SUBJ 10</u>	<u>SUBJ 10</u>
<ul style="list-style-type: none"> <li>• Steed</li> </ul>	<ul style="list-style-type: none"> <li>• The bulldozer (+</li> </ul>	<ul style="list-style-type: none"> <li>• A bowler hatted</li> </ul>	<ul style="list-style-type: none"> <li>• A bulldozer (+ driver)</li> </ul>

---

• Steed	driver)	man driving a car	• A bespectacled man
• Steed	• The evil man	• The man	• The second man
• Steed	• The evil man	• The first man	• The second man
• Steed		• The first man	

---

Table 1 shows that full NPs are maximally used 66% of the time, and minimally 14% of the time in narratives which are written in almost identical circumstances. Pronouns are minimally used 15% of the time to a maximum of 75%. Zero anaphors occur with a minimum percentage of 5% to a maximum of 40%; and this measurement pertains to only the two salient characters of the story. These figures show that narrative data collected through the procedure described above provide an extreme image of stylistic freedom. The point is, there is 'little homogeneity across subjects' produced narrative data. An ideal environment for investigating the cognitive constraints on Referential Choice needs the implementation of a method which does not allow the inconsistency existing in the data collected through a narrative production task to interfere with the results. If there is diversity across subjects' performances, it could be possible for us to show precisely what the possible causes are.

## 5. Discussion

The figures in table 1 provide a quite radical view of stylistic freedom. The longer the story and the fewer the referents introduced, the more established the protagonists become, consequently a greater number of attenuated markings are expected. The frequency of reference to the major characters of the story is certainly related to the possibility of using a greater number of pronouns. For example, in the case of subject eight, the frequency of 62% for full NPs indicates that the characters were not mentioned as many times as they could be, so fewer predications were made there. The differences in the frequency of full NPs and pronouns are caused by several factors, chief among them are: a) the variable length of the narratives; b) the differential number and frequency of entities introduced; c) initial introduction of referents with definite NPs as opposed to indefinite NPs; d) creating the appropriate semantic and syntactic environment for using particular constructions (e.g. the use of ellipsis depends on creating the appropriate semantic and syntactic environment); and e) greater elaboration of interactive parts, involving frequent subject switches and as a result more ambiguity is to be remove. These are the major variables causing variance in the referential patterns displayed in the subjects' written narratives.

The examination of the ten written stories demonstrate that: a) greater elaboration of the interactive part of the story involves more frequent use of full NPs which causes subject switches; b) the number of entities introduced and the frequency of the subsequent mentions of the entities introduced are correlated with the frequency of full NPs as opposed to pronouns used; c) characters initially introduced by an indefinite NPs require a different referential strategy throughout the story than characters introduced by definite NPs; and d) inanimate characters are not pronominalized, despite being back grounded information.

Diversity among discourse referential patterns might be caused by two types of factors: referential strategies of text producers and freedom of choice. The environments in which referential freedom is observed are limited: a) optionality in marking minor discourse breaks where there is no ambiguity involved; b) whether or not the presence of an expression is purely stylistic rather than being informative gives the opportunity to the text producer to decide upon the Referential Choice considering his desire; c) the interplay of conflicting factors makes the Referential Choice unpredictable (recency of prior mention, focus and referential salience within episodes lead to the use of an attenuated form; potential ambiguity, rhemic complexity, long referential distance and episodic breaks require producers to use a nominal form) and d) inappropriate choice of referring expressions may be made by less skilled speakers who do not help their addressee by making focus switches clear.

## 6. Conclusion

However, referential diversity may be caused by subjects' stylistic strategies which include factors such as: a) varying length of narratives produced; b) the number and the frequency of mentions of characters introduced; c) lengthy elaboration of different parts of the story; d) having the skill to use particular strategies which can affect the frequency of referential forms employed, e.g. the use of English ellipsis requires the writers to develop that skill; e) initial introduction of salient characters by a definite NP as opposed to an indefinite NP; and f) the number and frequency of insignificant entities introduced which contribute to greater complexity of the referential material can may result in the use of a full NP. The influence of individual's choice of strategies in



narrating a story results in diversity of referential patterns; but this type of diversity does not have anything to do with the effectiveness of discourse constraints.

To sum up, while a comparison of the individually written narratives can lead us to some prototypical results, it is hardly possible to reach conclusions about how individuals compare in their use of referring expressions in the discourses that they produce. It is observed that the function of definite and indefinite NP's in marking salient character is significantly different; the impact of such a difference is not limited only to the strategy taken in the first introduction. Saliency, animacy, pronominalisability and focus are issues which do not necessarily always agree. Salient characters are mostly animate and, when in focus, are pronominalizable. But inanimate referents may be less pronominalizable, e.g. quarry in the narrative data analyzed above is not pronominalizable because it is not animate; however, in certain contextual environments, an inanimate entity may be the center of focus and be frequently pronominalized, e.g. think of the ball in the soccer match. Indeed, pronominalisability doesn't ensure focus and saliency and non-pronominalisability also does not always indicate lack of saliency and focus.

There are many factors to explain the variability in the percentages of the use of referring expressions in this type of task: length of narratives, the writer's decision on which part to elaborate, and individual linguistic skills. The frequent use of ellipsis demands that the writers produce the semantic and syntactic environment for its use. Selection of different verb phrases impacts upon the frequency of mentions of referents, and the different possibilities for the number and frequency of mentions of characters all cause the diversity possible among individually written stories. We demonstrated that referential stylistic freedom and differing referential strategies may both be factors causing referential diversity; however, these do not interfere with the correlation between the choice of anaphoric expressions and the accessibility of referents marked.

## References

- Allan, K. (1986a). *Linguistic Meaning vol.1*. London and New York: Routledge and Kegan Paul.
- Allan, K. (1986b). *Linguistic Meaning vol.2*. London and New York: Routledge and Kegan Paul.
- Allan, K. (1995). *The analysis of discourse. Unpublished manuscript*. Monash University: Linguistics Dept.
- Almor, A. (2000). Constraints and mechanisms in theories of anaphor processing. In M. Pickering, C. Clifton, & M. Crocker (Eds.), *Architectures and Mechanisms for Language Processing*. Cambridge University Press. England.
- Ariel, M. (1985). The discourse functions of given information. *Theoretical Linguistics*, 12(2/3), 99-113. <https://doi.org/10.1515/thli.1985.12.2-3.99>
- Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, 24(1), 65-87. <https://doi.org/10.1017/S0022226700011567>
- Ariel, M. (1990). *Assessing NP antecedents*. London: Routledge.
- Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of Pragmatics*, 16, 443-463. [https://doi.org/10.1016/0378-2166\(91\)90136-L](https://doi.org/10.1016/0378-2166(91)90136-L)
- Botley & McEnery, T. (Eds.) (2000). *New Approaches to Discourse Anaphora*, Technical Papers.
- Chafe, W. L. (1980). *The pear stories: cognitive, cultural and linguistic aspects of narrative production*. Norwood, New Jersey: Ablex.
- Chafe, W. (1994). *Discourse, Consciousness, and Time. The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.
- Clancy, P. (1980). *Referential Choice in English and Japanese narrative discourse*.
- Clark, H. H., & Sengul, C. J. (1979). In search of referents for nouns and pronouns. *Memory and Cognition*, 7(1), 35-41. <https://doi.org/10.3758/BF03196932>
- Clifton, C. Jr., & Ferreira, F. (1987). Discourse Structure and Anaphora: Some Experimental Results. In M. Coltheart (Ed.), *Attention and Performance XII*. Hove: Erlbaum.
- Cornish, F. (1999). *Anaphora, Discourse, and Understanding. Evidence from English and French*. Cambridge: Cambridge University Press.
- Dale, R. (1992). *Generating Referring Expressions*. Cambridge, MA: MIT Press
- Gernsbacher, M. (1989). Mechanisms that improve referential access. *Cognition*, 32, 99-156. [https://doi.org/10.1016/0010-0277\(89\)90001-2](https://doi.org/10.1016/0010-0277(89)90001-2)

- Givon, T. (1983). Topic Continuity in spoken English. In T. Givon (Ed.) *Topic Continuity Discourse: A quantitative cross-language study*. Philadelphia: John Benjamins publishing company. <https://doi.org/10.1075/tsl.3.08giv>
- Givon, T. (1984 & 1990). *Syntax: A functional-typological introduction. Vols 1 & 2*. Amsterdam: John Benjamins publishing company.
- Givon, T. (1989). *Mind, Code and context: Essays in pragmatics*. Hillsdale, New Jersey: Lawrence Elbram Associates.
- Gordon, P., Grosz, B., & Gilliom, L. (1993). Pronouns, Names, and the Centering of Attention in Discourse. *Cognitive Science*, 17, 311-347. [https://doi.org/10.1207/s15516709cog1703\\_1](https://doi.org/10.1207/s15516709cog1703_1)
- Grüning, A., & Kibrik, A. (2002). Modelling Referential Choice in Discourse: A Cognitive Calculative Approach and a Neural Network Approach.
- Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, 69, 274-307. <https://doi.org/10.2307/416535>
- Kibrik, A. A. (1991). Maintenance of Reference in Sentence and Discourse. In W. P. Lehmann, & H. J. Hewitt (Eds.) *Language Typology*. (pp. 57-84). Amsterdam: John Benjamins. <https://doi.org/10.1075/cilt.81.06kib>
- Strube, M., & Wolters, M. (2000). A Probabilistic Genre-Independent Model of Pronominalization. Proc. of the 1st Meeting of the North American Chapter of the ACL. Seattle, WA, USA, April 29 - May 4, 2000, 18-25.
- Van Hoek, K. (1997). *Anaphora and Conceptual Structure*. Chicago: University of Chicago press
- Vonk, W., Hustinx, L., & Simons, W. (1992). The Use of Referential Expressions in Structuring Discourse. *Language and Cognitive Processes*, 7, 301-333. <https://doi.org/10.1080/01690969208409389>

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).