A Study of the Noun Phrase in Spoken and Written English

1 Introduction:
1.1 Aim and scope

The present paper is concerned with noun-headed phrases and pronoun-headed phrases in subject and object function in spoken and written English. The study is based on Samplers of the Christine corpus and the Susanne corpus in TIGERSearch software. Both are based on spoken and written English respectively. Sampler of the Christine corpus is a sampler of 200 sentences from the Christine corpus, which includes 200 numbers of corpus graphs and 895 number of tokens; on the other hand sampler of the Susanne corpus is a sampler of 200 sentences from the Susanne corpus, which includes 200 numbers of corpus graphs and 5257 number of tokens. Since both corpora are syntactic corpora with prominent tagging of noun-headed phrases and pronoun-headed phrases they facilitated my study in different texts with different functions. However, the corpora are so small that there might be a discrepancy between the results based on these two corpora and the results based on the big size corpora. So it is very necessary to confirm that my results are based on the present samplers only.

The aim of the paper is twofold: first, to investigate the frequency of pronoun-headed and noun-headed phrases in subject function and object function in spoken and written English; second, to compare the results with the present investigation with previous research.
1.2 Background

1.2.1 The noun phrase

According to Bakken (2006), the head of a noun phrase is a noun or a pronoun. If the head is a noun, it may combine with determiners, premodifiers and postmodifiers. On the contrary, if the head is a pronoun, determiners will normally not occur, and even if any modifiers occur these are usually post-modifiers. As the main elements of a sentence, the noun phrases may function as subject, object, subject and object predicative and prepositional complement. All of these mentioned above are nominal functions. In addition, noun phrases can function as adverbials, as vocatives, and as appositions. Furthermore, noun phrases can be used as an adjective to modify the head of noun phrase. A noun in the genitive may function as determiner in the noun phrase. Since noun phrase has different functions they are abundant in any text. If we take away the noun phrases from the text there is very little left.

Moreover, noun phrases are different in structures. Some of them are made up of single words while some of them are made of long complex structures. Normally, single words such as pronouns are used to clarify what we are referring to. If we want to express new information we normally need to use more complex noun phrases. According to Hasselgård et al. (1998) there are two types of noun phrases: one is noun-headed phrases; the other one is pronoun-headed phrase. Pronoun-headed phrases play different functions as the same as noun-headed phrases do. In addition, pronoun-headed phrases play different roles in the text. Excluding indefinite pronouns and wh-pronouns, pronouns can substitute full noun phrases when the referent is very clear in the situation or the surrounding text. Moreover pronouns greatly simplify the use of language, since they make it unnecessary to specify in detail who the speaker is, who the hearer is, and which other entities are being referred to. In another words, they are understandable and recognizable through the speech situation.

The following sentences are examples of noun phrases and pronouns in subject function and in object function:

1. we’ve # we’ve had it (pronoun as subject)
2. he makes jokes about me not being able to talk but I ain’t been able to talk (pronoun as object)
3. bound to be # bound to be awful when a woman can’t talk (noun phrase as subject)
4. He makes jokes about me not being able to talk but I ain’t been able to talk (noun phrase as object)

1.2.2 Some characteristics of speech and writing

Speech and writing are different in some aspects, which lead to different use of
noun-headed phrases and pronoun-headed phrases in different functions.

As we know, compared to writing speech is more informal in style. According to Hasselgård et al., (1998) informal speech differs from formal writing in a number of respects as follows:

First of all, informal speech is based on the face-to-face interaction. Normally, it is accompanied by body language. Since the speaker and the hearer(s) attend in the same situation they can often draw on a great deal of common background knowledge. So compared to writing speech tends to be less explicit. In speech, there is more situational reference, and its language is characterized by ellipsis and sentence fragments.

Secondly, informal speech normally has two or more participants who are in constant interplay. There is no such definite rule that they need to follow each other. That is why their conversation often overlaps. In speech, interrogative and imperative sentences are more common than in writing because they can attract the addressee(s) attention directly. Such as tag questions, and comment questions are used as interactive signals. According to them both speakers and hearers can interchange their thinking.

Thirdly, since informal speech is typically a social activity in speech speakers convey information less than in writing. For instance, in daily life people talk about the weather and use greeting words in order to communicate with their interlocutors. Of course, this kind of questions and greetings do not take much more information.

Fourthly, informal speech is produced in face-to-face environment. There is no possibility to edit the information which they are going to convey in the live situation. Moreover, listeners need to organize what they are going to talk about while they are accepting information which the speaker presented. On the contrary, speakers have enough time to conduct the information which they are processing and at the same prefer to accept the information from the listeners. Syntactically, spoken discourse therefore tends to be less complex, and there is less variation in both structure and vocabulary. There is a lot of simple colloquial phrases and more or less fixed collocations, which ease the on-line production of speech. Disfluency such as false starts, repetitions, and silent and filled pauses, are common.

In the end, there is a prominent difference between informal speech and writing. That is informal speech is based on the medium of sound. Of course, when we compare informal speech and writing we need to use the graphic system instead of the sound system. There is an obvious relationship between them. However, the sound system is far more changeable in the meanings it may express, especially through the system of intonation. In speech, according to intonation we can find the most important elements in the message, whereas, in writing we can find them in the
system of punctuation and from the organization of sentences, especially through the sentence structure. Moreover, intonation can express the emotional attitudes of the speaker in speech, but in writing even though there are few symbols such as italics, bold-face, capitalization to express the emotional attitudes of speaker and to emphasize the elements which are most important we find the means of expression are far more limited.

1.3 The corpus

1.3.1 background

Even though there are several definitions for the corpus, I prefer the definition which was given by Crystal (1991), namely "a collection of linguistic data, either written texts or a transcription of recorded speech, which can be used as a starting-point of linguistic description or as a means of verifying hypotheses about a language." Since corpora have the following three characteristics linguists tend to use corpora doing different kinds of researches: firstly, corpora are finite and representative samples of infinite language; secondly, corpora are edited (coded, annotated) for language studies, thirdly, Corpora can be text or speech (or both). A general aim of corpus linguistics is to check a language hypothesis against the actual use of language. Since there are limitations in the use of corpora we cannot draw conclusions directly from a corpus. What we first need to do is interpret the data using our linguistic intuitions and compare the results with the previous research. After then we can draw conclusions. On the other hand, a corpus may be useful to test our intuitions.

1.4 Hypotheses:

According to the different characteristics of spoken and written English which I mentioned in the prior passage, I think the frequency of pronoun-headed phrases in spoken language will be higher than in written when they function as a subject or as an object in the context. On the other hand, the frequency of noun-headed phrases in written language will be higher than in spoken when they function as a subject or an object in the context. Compared to the written English spoken English is more informal in style. In order to communicate well, in order to let the hearer get the main meanings from the conversation easily, in order to make the comfortable conversation situation it is very necessary to use simple sentences and informal sentences in spoken. However, in order to describe the things vividly and in detail the writer normally use the complex sentences with the heavy elements in written. Of course, there exist simple sentences in written as the same as in spoken. So my hypothesis is that pronoun-headed phrases in subject function are more frequent than noun-headed phrases in spoken, but pronoun-headed phrases in object function is less frequent than noun-headed phrases in object function in spoken English. On the other hand, noun-
headed phrases in subject function and object function are more frequent than pronoun-headed phrases in written English.

Furthermore, according to the principle of end focus the most important part of messages comes towards the end, while the least important parts come at the beginning of the clause. Normally, the least important part is what we know already. We also regard it as given information. However the most important part is new information. According to the principle of end weight the more weighty part of a structure should be placed towards the end. Most importantly, there is a close relation between the principle of end focus and the principle of end weight. Normally, given information can be described in few words since it is already known, while new information need to be expressed well. So my hypothesis to pronoun-headed phrases is that they will be more frequent in subject function than in object function both in spoken and written English. On the other hand, my hypothesis to noun-headed phrases is that they will be more frequent in object function than in subject function both in spoken and written English. Hence, my hypotheses are as follow: firstly, in spoken pronoun-headed phrases in subject function is more frequent than in object function, but noun-headed phrases in object function are more higher than in subject function. Secondly, in written pronoun-headed phrases in subject function are more frequent than in object function, but noun-headed phrases in subject function are less frequent than in object function.

2 Previous research

The use of pronoun-headed phrases and noun-headed phrases are different in syntactic function and in the type of text. According to Hasselgård et al. (1998: 141) “Pronouns are most typically associated with subject position; noun-headed phrases, with positions later in the clause.” They think that the reference which is known and familiar from the prior text normally occurs in the subject position. Because of this reason the reference can be expressed by a pronoun or a definite noun phrase. On the one hand, they think new information is generally introduced at the end of the clause by using the noun-headed phrases. This way to express the information is compatible with the information principle. Moreover, they think the use of pronoun-headed noun phrases and noun-headed phrases are different in the type of text. They think in conversation pronouns tend to outnumber noun and in academic prose and news reportage nouns are about ten times more common than pronouns with the more complex structure and more heavy information.

According to Bakken (2006), the new and most important part of information come towards the end of the structure. She thinks the concept of end focus and end weight are closely connected with the principle of given and new information. Given information is referred to something which is known, while new information is
presented as unknown. “…new information is often presented with longer, heavier structures, such as complex noun phrases, while given information tends to be expressed through the use of pronouns or other pro-forms.”

3 Methods:

My purpose is to investigate the frequency of pronoun-headed phrases and noun-headed phrases in the subject function and object function in different discourses, especially in spoken and written language. In present paper I used the corpora of TIGERSearch software suite, which is syntactically annotated. The reason for using TIGERSearch tool is as a tool TIGERSearch tool is for linguistic text exploration. Through TIGERSearch it is easy to find linguistically annotated texts and it is very easy to obtain sample sentences for the syntactic phenomena the investigator are interested in. In TIGERSearch software suite five groups of English samples can be found. What I am concerned in the paper is Chrissampler and Suesampler. In order to search the frequency of pronoun-headed phrases and noun-headed phrases in subject and object function in different discourses it is necessary to use TIGER query language to search, which can either be drawn in the graphical query mode or be written in text form mode. Of course, TIGER language is general description language for syntax graphs, which is used to express syntactic structure. According to Smith (2003), a tag for part of speech used in the TIGER Treebank is the Stuttgart-Tubinger Tagset, with minor differences. A tag for a word form consists of the tag for the main word category followed by tags for the subcategories. A search for wordforms belonging to a particular category is accomplished by using an expression in the TIGER language known as a node description. In TIGER language the simplest node make up of an expression known as a feature constraint. If it is the simplest feature it will consist of a single feature-value pair, a feature and a value are separated by an equal sign. Furthermore, constituent categories are encoded in node labels: non-terminal node labels represent phrasal categories; terminal nodes represent wordforms and are tagged. Moreover, in TIGER language syntactic functions are encoded in edge labels. In other words, in TIGERSearch format edge labels contain the original syntactic function tags and the (nonterminal) cat category contains phrase and clause forms. Since the graphical query editor consists of two regions: the word level ( at the bottom) and the nonterminal tree level ( on top) it is easy way to draw the graph and switch it into textual mode to get the searching formulae. That is actually the way how I get the following searching formulae:

1. [NT] >s #n2:[NT] & #n2 > [pos = /PP.*/] for searching pronoun-headed phrase as subject
2. \([NT] >o \#n2:[NT] \& \#n2 > [pos = /PP.*/] \) for searching pronoun-headed phrase as object

3. \([NT] >s \#n2:[NT] \& \#n2 > [pos= / N.*/] \) for searching noun-headed phrase as subject

4. \([NT] >o \#n2:[NT] \& \#n2 > [pos= / N.*/] \) for searching noun-headed phrase as object

In above formulae \(/ \) represents the node, \( > \) expresses relations of dominance, \( \#n2 \) refers to the same node, \( \& \) refers to “and”. \( NT \) refers to the nonterminal node and \( pos \) refers to part of speech. \( PP.* \) means starting with pp, which refers to all pronoun-headed phrases and \( N.* \) means starting with N, which refers to all noun-headed phrases. \( S \) and \( o \) refer to subject and object respectively and both of them are edge of TIGER Search query language.

4 Results

Table 1: The frequency of pronoun-headed phrases and noun-headed phrases in subject function and object function in spoken and written text

<table>
<thead>
<tr>
<th></th>
<th>Frequency In Spoken</th>
<th>Frequency In Written</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pronoun-headed phrase as subject (formula 1)</td>
<td>91</td>
<td>57</td>
</tr>
<tr>
<td>Pronoun-headed phrase as object (formula 2)</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Noun-headed Phrase as subject (formula 3)</td>
<td>10</td>
<td>150</td>
</tr>
<tr>
<td>Noun-headed Phrase as object (formula 4)</td>
<td>25</td>
<td>131</td>
</tr>
</tbody>
</table>

5 Comparison and Discussion:
5.1 Written and spoken language

According to the table 1 we can find that the frequency of pronoun-headed phrases in spoken (subject 91/ object 13) is higher than in written (subject 57/ object 10) when they function as subject or as object in the sentence. On the other hand, the frequency of noun-headed phrases in subject function (150) and in object function (131) in written is higher than in spoken (10 subject / 25 object). The results are compatible with the characteristics of speech and writing and previous research done by Bakken (2006) and Hasselgård et al. (1998).

5.2 Subject and object function in spoken English

According to Table 1 we can see the frequency of pronoun-headed phrases in subject function (91) is seven times as much as pronoun-headed phrases in object function (13) in spoken English whereas the frequency of noun-headed phrases in subject function (10) is smaller than in object function (25) in spoken English. These results are related to the special role of pronouns, the information principle (principle of end focus) and principle of end weight and the special type of context. As we know, pronouns can replace noun phrases of the former context and can refer back to the noun phrases in the previous context and make the sentence structures more simple and make the sentence more easy for the readers to accept. Meanwhile, according to the information principle of end focus and end weight principle, the most important and heavy elements occur at the end of sentences. Compared to the noun-headed phrases, pronoun-headed phrases are much more simple in structure than the noun-headed phrases. So it is easy to occur in the subject position at the beginning of the sentence. Since object position is at the end of the sentence, where the important part of speech occurs, which is normally new for the readers. Of course, noun-headed phrases have such functions to express new information. As we know, the head of the noun-headed phrases is a noun, which can be followed by the premodifiers and post-modifiers. Since the structure of the noun phrases are more complex than the pronouns they come at the end of the sentences.

5.3 Subject and object function in written English

According to the Table 1 we can find that the frequency of pronoun-headed phrases in subject function (57) is higher than in object function (10) in written English and the frequency of noun-headed phrases in subject function (150) is higher than in object function (131) in written English.

The use of pronoun-headed phrases and noun-headed phrases varies with syntactic function and with the type of text. No matter it is in spoken or written pronoun-headed phrases are most typically associated with subject position. It is well known that
initial position is the most important position since elements with a linking function are placed there. Such elements can show the relationship between different points in the context; thus they are capable of achieving cohesion in the text. As subject in a sentence typically makes reference to something which is known it can be expressed by a pronoun or a definite noun-headed phrase. So its frequency in subject function is higher than in object function. New information is generally introduced later in the clause, using noun-headed phrases (usually indefinite in form). Due to my view, the frequency of noun-headed phrases should be higher in object position than in subject position. However, it arrives different results in figure. This might be related to the size of the samples.

5.4 Pronoun-headed and noun-headed phrases in spoken English

According to the table 1 we can see in subject function the frequency of pronoun-headed phrases (91) is nine times as much as the frequency of noun-headed phrases (10) in spoken English. On the other hand, in object function the frequency of noun-headed phrases (25) is higher than the frequency of pronoun-headed phrases (13). As we see, the frequency of pronoun-headed phrases in subject function is much higher than the frequency of noun-headed phrases in spoken English. The reason is as already discussed that spoken English in structure is more easier than written English. Since both speaker and hearer are possible to occur in the same situation they are prone to use the pronoun to refer to someone or something which both of them are familiar with in spoken English. In object function noun-headed phrases (25) are more frequent than pronoun-headed phrases (13). The reason is that the most important information and the heaviest elements come at the end of the sentence. Noun phrases have such a capability to pack the most important and heavy information, and to transfer them to the listeners. So they tend to occur in the object position.

5.5 Pronoun-headed and noun-headed phrases in written English

As we see in table 1, the frequency of noun-headed phrases (150) in subject function is three times as much as the pronoun-headed phrases (57) in subject function in written English; the frequency of noun-headed phrases (131) in object function is twelve times as much as the pronoun-headed phrase (10) in object function in written English. From the results we can see that in written English the frequency of pronouns is not similar with in the spoken. To some degree it is related to the different types of discourses. As we know, the structure of written texts is more complex than the spoken. In order to describe something vividly or in detail the elements in written sentence normally are accompanied by the heavy structures. Noun-headed phrases
have such a kind of characteristics. However, pronouns are very simple in structure.

5.6 Subjects in spoken and written language

According to the Table 1 we can see the frequency of pronoun-headed phrases (91) in subject function in spoken language is higher than in written English (57), and the frequency of noun-headed phrases in subject function (150) in written English is much more higher than in spoken English (10). The reason is related to the characteristics of speech and writing. As I mentioned before, in spoken the choice of pronoun is the easy way to refer to something which is already mentioned. It is the best the choice when speakers don’t want to mention somebody whom they are talking about or whom they do not want to mention. Moreover, spoken English structure is simpler than written English. On the one hand, in subject function noun-headed phrases (150) are more frequent in written English than in spoken English (10). The reason is that noun-headed phrases are capable to take the most important information and new informations according to adding the premodifier and post modifier. In addition, writing is different from speech in the environment of expressing the information. As I mentioned, in speech speakers and listeners participate in the same environment and they can get the common background for the conversation. However, in writing the author needs to describe the things in detail and vividly to get the readers’ reaction. The reasons above make the results different from the former.

5.7 Objects in spoken and written language

According to the Table 1 we can see the frequency of pronoun-headed phrases in object function (13) in spoken language is a little bit higher than in written language (10), and the frequency of noun-headed phrases in object function (131) in written is five times as much as in spoken English (25). According to the information principle and principle of end weight it is acceptable that in written in the object position there occurs a lot of noun-headed phrases. However, we can find that the frequency of pronoun-headed phrases are not much even in the spoken. To some degree, it is related to the position of the object. Since it appears at the end of the clause it is easy to take the heavy elements and most important elements. In addition, noun-headed phrases have such a kind of ability to express the information.

6 Conclusion:

According to the sampler of Christine corpus and the sampler of Susanne in TIGERSearch software the frequency of pronoun-headed phrases in subject function in spoken and written English is much higher than in object function. The frequency
of noun-headed phrases in subject function in spoken is smaller than in object function, but the frequency of noun-headed phrases in subject function in written is much higher than in object function. Compared to the noun-headed phrases, pronoun-headed phrases in subject function are more frequent in spoken English, but pronoun-headed phrases in object function have small frequency in spoken English. On the other hand, in written the frequency of noun-headed phrases in subject function and in object function is much more higher than pronoun-headed phrases. Generally, pronoun-headed phrases in subject function and in object function in spoken is higher than in written English; whereas noun-headed phrases in subject function and in object function in written is much higher than in spoken English. Of course, there is one thing I need to stress that the conclusion I draw in my present paper is based on two small corpora.

REFERENCES

http://www.ims.uni-stuttgart.de/projekte/TIGER/TIGERSearch/papers.shtml


