Style and Readability in Football Texts
# TABLE OF CONTENTS

1. INTRODUCTION 3

2. THEORETICAL PREMISES 5
   2.1 Style 5
       2.1.1 Style and Register 8
       2.1.2 Different Genres and Different Styles 10
       2.1.3 Readability 12
   2.2 Overview of Stylistics 14
       2.2.1 A Brief History of Stylistics 16
       2.2.2 Current Trends in Stylistics 18
   2.3 Style and Stylistics in This Study 20
   2.4 Languages for Special Purposes 21
       2.4.1 The Language of Football as an LSP 25

3. PREVIOUS RESEARCH 27

4. MATERIAL AND METHOD 30
   4.1 Quantitative Test 34
       4.1.1 Problems with English Participles 35
       4.1.2 Problems with Finnish Participles 36
       4.1.3 Other Problems Related to Categorization 37
   4.2 Qualitative Test (Questionnaire) 37

5. ANALYSIS 39
   5.1 Quantitative Analysis and Its Results 39
5.1.1 Average Sentence Length
5.1.2 Nouns
5.1.3 Verbs
5.1.4 Pronouns
5.1.5 Adjectives
5.1.6 Numerals
5.2 Expectancy in Readability
5.3 Qualitative Analysis and Its Results
5.4 Discussion of the Results of the Qualitative Analysis
   5.4.1 The Effect of Previous Knowledge on Readability
5.5 Comparison between the Methods

6. CONCLUSION
REFERENCES
APPENDICES
1. INTRODUCTION

This master’s thesis deals with the language of football, i.e. soccer, as a special language. In particular, it is concerned with its style and readability. Choosing texts about football as research material was motivated first and foremost by my long-time interest in the game. As far as I know, the style and readability of football-related texts have not been studied before.

In the analysis, style and readability are first approached from a mathematical point of view by paying attention to quantitative properties, such as sentence length, and the frequencies of nouns, adjectives and verbs. Secondly, the style and readability of the chosen research material are tested with readers by using a questionnaire.

Because a quantitative approach would, however, hardly lead to satisfactory results about the overall style and readability of the analyzed texts, I apply a qualitative method, i.e. the questionnaire, as well. The appeal of quantitative analysis is, at least in the present case, based largely on a sense of objectivity. This is also the very reason why I do not analyze the qualitative properties of the material myself, but subject it to a reading test taken by a group of readers. Hence, I hope to achieve as objective results as possible. The material and the methods are introduced in more detail in chapter 4.
The specific research questions that I seek to answer are as follows:

1) Are there any significant differences between football texts representing special language and football texts written in general language (GL)?

2) Are there any significant differences between Finnish and English football texts that fall into the same category as far as text type is concerned? If so, can this information be of help when translating?

3) Do the results provided by the quantitative and qualitative methods support each other, or is there a contradiction?

4) If the language of football is seen as a language for special purposes (LSP), how does it compare to other such languages?

Question 2 needs some clarification. Since the corpus does not consist of original texts and their translations, possible implications on translation are largely based on those results received in this study that seem to be in violation of what could be regarded as good readability. In order to answer the foregoing questions, it is necessary to introduce the theoretical premises relevant to this study.
2. THEORETICAL PREMISES

This chapter deals first and foremost with the concepts of style, register, genre and readability, and the connection between them. Section 2.2 provides a brief overview of the field of stylistics, and section 2.3 focuses on definitions of special languages. Finally, there is a discussion on how these definitions can be used to show that the language of football can be considered a special language.

2.1 Style

It is quite often that you hear people use expressions such as the Dickensian style to refer either to the qualities of a specific writer or writers that are seen as belonging to the same group as the writer mentioned in the expression. There is nothing wrong with such expressions, but it would be misguided to believe that the style used in “Dickensian” novels would be exactly similar to that of Charles Dickens. In other words, the term Dickensian can be used when there are one or more characteristics reminiscent of Dickens’ style. In truth, it may be exceedingly difficult to find such common stylistic qualities even when comparing books by the same author – let alone when comparing the works of several authors that are supposed to represent the same style. (Leech & Short 1987: 11–12.)
However, it is not by chance that people write texts the way they do. To put it simply, style is always a conscious choice dictated by ideas people form when approaching different kinds of texts. People adhere to widely accepted stylistic regularities, *norms*. It is this adherence that in turn leads to the emergence of different styles typical of specific spheres, such as science. (Saukkonen 1984: 9.) Furthermore, this is connected with different *genres*, such as poems, business letters, newspaper articles, and so on. Genres can be described as texts used in a particular situation for a particular purpose, and they have features which differ from those of other genres (Trosborg 1997: 6) – a phenomenon that will be dealt with in section 2.1.2.

Even though people are somewhat bound by norms, it does not mean that there is no room for personal styles. Different kinds of texts limit our stylistic freedom to different extents. In this sense, a research paper is quite restrictive whereas a poem offers virtually countless possibilities as far as style is concerned. Saukkonen (1984: 11) points out that it is not always easy to put oneself in the reader’s position and see the text through his or her eyes. Therefore, official and professional texts are always written in a specific style to ensure that communication is successful.

Traditionally, style is seen as belonging mainly to the domain of written literary texts (Leech & Short 1987: 11). Literary texts are texts in which single words have an important role: they are something that the reader can enjoy. Further, they are meant to appeal to the reader’s imagination. Drama, fiction and comedy are good examples of such texts. By contrast, non-literary texts belong to the domain of reality and fact.
This is to say that their content is always more important than their form, as is the case, for instance, in encyclopedias. (Newmark 2004.)

Saukkonen (1984: 29) considers a trichotomy between artistic, informative and scientific-theoretical styles to be most apt to function as the main categorization of the styles of today when historical reality is used as a standpoint. The purpose of artistic texts is to create a specific atmosphere, feeling etc. that the readers can understand any way they like. In other words, there is always room for interpretation in artistic texts. Scientific texts, on the other hand, aim to do the exact opposite. There should not be room for different kinds of associations; rather, the texts should be as unambiguous as possible. Further, scientific texts do not appeal to the reader’s emotions, but to reason, intelligence and special knowledge. (Saukkonen 1984: 36.) Artistic style abounds in expressions such as cat’s paw, whereas scientific texts would, in this case, probably refer to a gullible victim. The use of metaphorical language leads to the “distortion” of reality because the reader cannot take the text literally. Formal style, which is used in scientific texts, is to be taken literally, as it should contain no connotations (Rainio 1981: 14).

Informative style can be placed between artistic and scientific-theoretical styles, in that it usually aims to be as precise and “formal” as scientific-theoretical style, but it is also based on common knowledge – a feature shared with artistic style. Informative texts are also often channels where writers may express their opinions and seek to affect the reader in some way. Hence, it is evident that all three styles have different
functions, as they approach their topics from different angles. Scientific texts depict reality in analytical and artistic texts in a metaphorical way. Informative texts, on the other hand, may generalize without being scientifically analytic, or simply describe reality “as is”, which can never be done in scientific texts. These differences in angle, in turn, make scientific style quite abstract whereas artistic style, by contrast, appears as concrete. Again, informative style is somewhere between these two. (Saukkonen 1984: 36–38.)

Saukkonen (1984: 39–40) finds that artistic style is more dynamic than scientific style. This is to say that verbs in artistic texts tend to be more concrete than those in scientific texts. For example, push, pull, throw etc. may be easily found in artistic texts while such verbs as divide or categorize are frequent in scientific texts. Even though there are a number of other features that could be used to further distinguish between the three styles, the already mentioned features should be sufficient to help explain the primary differences between the texts that will be analyzed in this study.

2.1.1 Style and Register

A concept closely related to genre is register, a linguistic subset for a specific purpose or a social setting. Probably one of the most evasive terms in linguistics, register can be seen as the framework of style. Suitable frameworks, or appropriate registers, are in turn defined by genres. Trosborg (1997: 5–7) states that registers form a group of varieties, i.e. styles typical of professional fields, such as religion,
law etc., and can therefore be contrasted with user-specific geographical, temporal and social varieties. For instance, legal register comprises the language of different kinds of legal documents be they wills or contracts. Sometimes, it is possible for genre and register to virtually overlap, as is the case, among other things, in weather forecasts because the genre as a whole is quite restricted. This, however, does not happen very often, as register is a purely linguistic feature whereas genre is tied to the function of texts in society, the defining criterion of which is its communicative purpose (Trosborg 1997: 8–9).

Knowing where and when to use different registers is of utmost importance if people want to succeed in communicating their ideas: a scientific article written in a poetic way would not be understandable, or would not be understood in the right way. It is perfectly suitable to use dialect in a novel, but not, for instance, in a job application. This is to say that different contexts have different norms, and different audiences have different requirements and expectations. (Saukkonen 1984: 10.) What we intend to say is also a factor that affects both style and register. Apologizing, for instance, is usually done in a formal manner, whereas telling a joke does not have this requirement. A failure to recognize the style or register required in a certain situation may result in confusion, criticism or even embarrassment. (Crystal & Davy 1973: 4–5.)

As already implied above, idiolects, personal languages and styles, are affected by several factors. Our style is not influenced only by our inner world, but also by the
world around us. For instance, newspaper articles written, say, a hundred years ago are stylistically quite different from those written today, and what is considered as formal style in Finland may be considered inappropriate in some other country. In addition, social status is likely to affect personal style – an erudite person knows how to write formal texts whereas an uneducated individual may have difficulties with finding the right register.

2.1.2 Different Genres and Different Styles

Style and genre are so closely connected that there cannot be one without the other. Whenever people follow stylistic norms, they do so to fulfill the criteria of a given genre. As mentioned in section 2.1.1, genres are tied to the function of texts and they are defined by their communicative purpose (Trosborg 1997: 8–9). Adhering to norms is simply a means to make communication more understandable.

Different genres and styles therewith have come into being because of international cultural development. In the past, the use of language was more homogeneous, albeit different styles, such as poetic and informal styles, already existed. As linguistic communities started to become more and more specialized and complex in nature, so did language, and new styles and genres began to emerge. (Saukkonen 1984: 14.) During the past decades, such development has been more rapid than ever before, as mass media, TV, radio and the Internet have affected the way people communicate all
over the world. Differences between genres are of considerable importance in this study and will therefore be dealt with next.

Sanders (1977) divides style into three main categories that are *Gebrauchssprache* (formal style), *Kunstsprache* (artistic style) and *Alltagssprache* (informal style). The first category consists of texts written in the form of formal prose. Hence, scientific style, official style and the style of the press fall into this category. Artistic style, in turn, covers epics (novels, short stories, fables, fairytales etc.), lyrics (odes, hymns, ballads, elegies etc.) and dramatics (tragedies, comedies, historical dramas, operas etc.). In informal style, communication and contact are a goal in their own right, but in formal style, communication is a way of informing the reader of something while artistic style is a combination of both communication and aesthetic creativity. The form of the text is most evident in works written in artistic style and least evident in texts written in formal style.

According to Saukkonen (1984: 27), texts can be divided typologically into three main categories based on universally logical principles and independent of cultural traditions. It is common to use grammatical “persons” as the basis of this kind of categorization:

1) expressive texts focused on the speaker

2) directive and pleading texts focused on the addressee

3) declarative and descriptive texts focused on the topic.
One should bear in mind, however, that these categories are not descriptions of any particular genres, as they are rather artificial, abstract and formal in nature. The main categories can further be divided into virtually countless subcategories that can but do not have to be culturally bound. (ibid.) For instance, expressive texts include scientific texts, directive texts manuals and text books, pleading texts adverts, and declarative and descriptive texts stories, just to name a few.

However, one might wonder whether it would be possible for one text to be, say, both expressive and speaker-oriented, and descriptive and topic-oriented. Indeed, a single text may have all these qualities, and this is also noticed by Reiss (1989: 105) who divides texts into expressive, informative and operative types, but places, for instance, biographies somewhere between expressive and informative text types. Reiss’s categorization will not be introduced in more detail here, as I consider the model introduced by Saukkonen (see discussion above) to be drawing on the former.

2.1.3 Readability

Readability is yet another characteristic of texts inevitably affected by the writer’s stylistic choices. If a text is easy to read, it is readable but there are, of course, different levels of readability. These levels are determined by the structures used in texts, word choices, as well as the appearance of texts. If the reader is familiar with the subject dealt with in the text, it is naturally easier for him or her to read the text.
Hence, it is no wonder that readability and understandability are often seen as synonyms. (Hirsjärvi, Remes & Sajavaara 2000: 261–262.)

Savolainen (2003; the example in the parentheses by the author) lists a few examples of structures that may make reading difficult:

* Long and complex sentences that fail to express ideas in a coherent way.
* Excessive and unnecessary use of nouns.
* Excessive use of determiners with nouns.
* Clauses separating parts of the main sentence from each other in a way that may force the reader to read the sentence again (e.g. The man, who we met for the first time yesterday and had breakfast with this morning, is a thief.).

In addition, long words, and the use of the passive voice instead of the active may also have a negative effect on readability. If a text contains a great number of numerals, it may make the text harder to understand and to remember. According to Francis and Kučera (1982: 546), a high frequency of pronouns is a sign of anaphoric style, which means that a pronoun is used instead of repeating a word used earlier. Pronouns may also make texts more cohesive, thus improving their readability. Obviously, all these factors also affect reading speed. The faster a text can be read and understood, the more readable it can be expected to be. It should be borne in mind, however, that it is not always possible, or even wise, to avoid all the structures listed above.
Readability has been studied since the 1920s, and during this time, researchers have developed many formulae to measure the readability of texts. The earliest formulae were mathematical and focused mainly on the length of both words and sentences. Because the results provided by these methods proved somewhat unreliable, researchers had to come up with more complex tests. This resulted, among other things, in the development of cloze tests, reading speed tests and questionnaires that measure how well the readers have understood the text and how well they remember it. (Virtaluoto & Väyrynen 2000: 1, 3.) However, as Crystal (1997: 254) points out, the use of these more modern methods is quite cumbersome, and that is why some of the old formulae are still widely used.

2.2 Overview of Stylistics

Stylistics can be defined as research into style, and it is needed for two essential reasons. First of all, people need to know how and when to use different varieties of language. In other words, they have to adapt to all kinds of linguistic situations. This is naturally of great importance especially to students of foreign languages, but also to native speakers, as people in different walks of life, for example, have their own language and jargon. Secondly, stylistics may offer the means to resolve problems related to linguistic expressions in the vein of “is this expression adequate or effective enough?”. However, this does not imply that one should not take his or her intuitions into account. Stylistics merely offers an alternative approach that can be used to
justify different opinions. In brief, stylistics may help to find out where ambiguities and irregularities lie. (Crystal & Davy 1973: 6–9.)

According to Crystal and Davy (1973: 10–11), the aim of stylistics is first and foremost to identify and analyze linguistic elements that stand out from the mass of common, everyday language as being restricted to a specific social context. It is also the task of an analyst to try to explain why and to what end such stylistically significant features have been used, and why they have been used instead of other alternatives. This last point is quintessential to stylistics – after all, stylistic analysis is largely based on comparison.

The first phase in stylistic analysis is always an intuitive one. In this sense, stylisticians are equal with laymen, as they all see features they consider stylistically significant. The difference between the two groups is that stylisticians have developed a stylistic eye for themselves. In other words, it is easier for them to notice features that are likely to be stylistically important. More importantly, stylisticians know how to analyze their findings by using their knowledge of different theories. To put it simply, the main difference between a stylistician and a layman is that the former is a linguist whereas the latter is not. Because of their expertise, stylisticians are able to distance themselves from their personal opinions and to talk about language objectively. (Crystal & Davy 1973: 12.)
It is difficult to say whether stylistics forms a scientific field of its own. It could be seen as part of linguistics or literary science if the methods applied are typical of either one of these fields. This, however, would narrow down the scope of stylistics too much, especially if the subject of analysis happens to be as complex and versatile as written texts. On the other hand, seeing stylistics as a field of its own leads to problems as well. This is mainly because the study of literary language, literary texts, and literary discourse has not resulted in the manifestation of theories or methodologies especially typical of stylistics. (Leppänen 2000: 221.) What is more, the fact that literary studies are always to do with style makes stylistics as a term that much more elusive (Bradford 1997: xi).

Stylistics is not even a standard term, which is also a good indication of the multifaceted nature of the field. Stylistics has many subcategories, such as literary semantics, literary semiotics and literary rhetoric. (Leppänen 2000: 221.)

2.2.1 A Brief History of Stylistics

The roots of stylistics are deep in history. Bradford (1997: 3) situates them to fifth century Sicily where Greek settlers started to study and practise rhetoric. Nonetheless, it was not until the 1920s and 1930s that two major movements, Formalism and New Criticism, emerged. The goal of both the movements was the same – “to define literature as a discourse and art form and to establish its function as something that can be properly studied”, as Bradford (1997: 12) puts it.
In spite of this, Formalists and New Critics worked separately, the former in Europe and the latter in Great Britain and America. This was the situation until the 1960s when the movements recognized their similar goals, but, by then, new movements, theories and methods had already emerged and challenged the old ones. These included, among others, structuralism and feminism, but also other movements that stemmed from outside the fields of rhetoric and aesthetics. (Bradford 1997: 12–13.)

When discussing the early stages of modern stylistics, many scholars (see e.g. Bradford 1997: 36; Leppänen 2000: 219–220) tend to refer to Roman Jakobson’s famous closing statement at the Style in Language Conference held in Indiana in 1958. According to Leppänen (2000:219), the following words are said to have brought literary language into the scope of western linguistics:

> If there are some critics who still doubt the competence of linguistics to embrace the field of poetics, I privately believe that the poetic incompetence of some bigoted linguists has been mistaken for an inadequacy of the linguistic science itself. All of us here, however, definitely realize that a linguist deaf to the poetic function of language and a literary scholar indifferent to linguistic problems and unconversant with linguistic methods are equally flagrant anachronisms. (Jakobson 1960: 377.)

This is the background from where modern stylistics came into being. According to Bradford (1997: 13–14), modern stylistics is divided into two different kinds of approaches, textualist and contextualist. Formalists and New Critics favour the textualist approach, a method that focuses only on the text and nothing else that could
be related to it, such as the writer’s intentions. These two groups still cherish the belief in the difference between literature and other forms of language, and seek to specify it. From their point of view, literary style does not belong solely to the domain of literature but to all language. Text is considered literature only when it is dominated by combinations of stylistic features. In contrast, the methods and theories applied by contextualists tend to be less restricted, and as one could assume, they are mainly focused on the relation between text and context.

2.2.2 Current Trends in Stylistics

The trends in modern stylistics emphasize contextualism, and therefore, also pragmatic aspects. In other words, the object of analysis is the relation between the reader and the text in a given historical context. The reader’s knowledge and the associations that he or she makes, as well as the place where the interaction between the reader and the text occurs, are all of particular interest to the present-day stylistician (Leppänen 2000: 239–240; Ashurova 2006). According to Leppänen (2000: 240–241), however, text-oriented approaches are still quite common, which can be regarded as a remnant of New Criticism, or a feature typical of ideas about texts, reading and understanding in western culture.

Style, as a whole, is nowadays seen as a social and cultural phenomenon the nature of which is defined by the very community interpreting it. After all, style is a phenomenon affected by many extralinguistic factors, such as time, culture, politics,
literary traditions etc. (Leppänen 2000: 242; Ashurova 2006). This opens new and interesting viewpoints into culture and societies, as well as values and ideologies. Hence, it is secondary to ask what literary texts mean – it is more important to know how and why they mean what they mean. (Leppänen 2000: 243.)

Leppänen (2000: 244) points out that scholars have demanded that literary stylistics be more credible and repeatable, but at the same time, it has been suggested that literary stylistics should not aim at being neutral and objective. Bearing in mind that many literary texts are ambiguous on purpose, interpreting them in an objective and truthful way is close to impossible. Therefore, it is necessary to consider what the position of the researcher is. As Leppänen (ibid.) points out, this is the very standpoint, for instance, in critical and feministic stylistics, and they do not even try to come up with objective results, but, rather, to challenge old methods.

So, today, stylistics seems to be quite a heterogeneous field. Different schools have adopted different influences, methodologies, tools and concepts from various fields, such as linguistics, psychology, sociology and cultural studies. For instance, the latter has widened the scope of stylistics into popular literature, comics and scientific texts. Even though this kind of selection of tools has, without a doubt, enriched the field as a whole, it has also led to problems. It may be outright impossible for different schools to exchange ideas in a way that would benefit all parties, as there are no common terminologies or methods. What is more, even the objects of study are something that scholars cannot always agree upon. Even so, stylisticians have given
much to linguistics, as they have, for instance, developed and tested linguistic methods, and shown the significance of literary texts in respect of the teaching of languages. (Leppänen 2000: 236–238.)

2.3 Style and Stylistics in This Study

As the reader will have learnt by now, style and stylistics are not easy to define. After all, everybody has their own distinct style and opinions about style. One of the few conclusive remarks that can be made about style is that it seems to be more or less bound by norms in different genres (see discussion in section 2.1) which nevertheless leave room for personal aspects of style.

Style has traditionally been seen as belonging to the domain of literary texts (Leech & Short 1987: 11; see section 2.1 in this study), but this view is abandoned in the present study, as the research material also contains extracts that are purely informative in nature. Hence, it is possible to see style through two different classifications presented by Saukkonen (see section 2.1), one of them being a trichotomy between artistic, informative and scientific-theoretical styles, and the other a typological categorization of texts into the following groups:

1) expressive texts focused on the speaker
2) directive and pleading texts focused on the addressee
3) declarative and descriptive texts focused on the topic.
Stylistics is a heterogeneous field that has not managed to establish its position as an independent discipline. As mentioned above, there are countless different opinions about style, so it is no wonder that style has been studied from many different angles over the years. Even though a number of different viewpoints and theories are presented in sections 2.2.1 and 2.2.2, they merely function as a point of comparison to better explain the multifaceted nature of stylistics.

The present aim, however, is merely to follow the primary principle of stylistics: to identify and analyze linguistic elements that stand out from the mass. Here, a quantitative approach, i.e. counting the frequencies of grammatical elements, is combined with a qualitative one. In order to better maintain a sense of objectivity, the latter is realized as a questionnaire, which some could also see as containing quantitative features. Hence, the present study represents quantitative stylistics for most parts. Now we shall turn to yet another important concept for this study: languages for special purposes, or LSPs for short.

### 2.4 Languages for Special Purposes

Languages for special purposes are languages used in different kinds of scientific or professional fields, and – at least in some cases – when talking about a hobby. LSPs are grammatically quite similar to general language (GL), but they may be hard to understand nonetheless. This is mainly due to unfamiliar vocabularies, terms and
concepts. Sentence structures in LSPs do not usually make understanding difficult, but this is often the case, for instance, in legal documents where they do not, however, serve any necessary communicative purpose. (Haarala 1981: 9; Gustafsson 1996: 39, 42; Yli-Jokipii 2004: 82–84.) According to Gustafsson (1996: 39), the same applies also to many contractual texts, for instance, maritime documents. Varieties of language of this kind are considered so hard to understand that people have actually started to campaign against them, the most notable example of this probably being the Plain English movements of the USA and England dating to the late 1970s. (Crystal 1997: 382.)

Technical language, as opposed to the examples above, tends to use long and heavy noun phrases even though the sentences themselves are quite simple, as, for instance, Varantola (1984) has shown. Technical language is therefore probably a better example of a “typical” LSP. After all, it has even been suggested that there are no LSPs, only special vocabularies (Sager, Dungworth & McDonald 1980: 2).

According to Haarala (1981: 10), it is not easy to say where an LSP ends and another one begins. This is to say that LSPs have common terms, such as offside, used, among other things, in football and ice hockey. The same applies to the relation between LSPs and GLs. The word mouse, for example, has its own meaning both in GLs and when referring to computers in LSPs, where it is a term. In short, LSPs do not overlap only with each other, but also with GLs.
Interestingly enough, though, scholars have started to become interested in the changes of terms and concepts only during the last few decades, which is largely due to the emergence of socioterminology, a new branch of terminology. Indeed, recent studies have shown that LSPs are anything but stagnant: they change as the world and our understanding changes. (Pasanen 2003: 244.)

LSPs must meet certain requirements to serve their purpose. First of all, they have to be unambiguous, accurate and logical. In addition, they need to follow the norms set for good GL, which means that they are easy to use and linguistically correct (SK: 12). To further distinguish between GLs and LSPs, one has to move away from grammatical features and take other factors into account. The most important factor of this kind is probably the topic concerned, as LSPs are usually considered “special” mostly because of their content. In other words, the content does not consist of general knowledge, but it is dictated by the writers’ knowledge of the special field and what they expect the reader to know. (Sager 1994: 40)

LSPs are usually meant for communication between professionals, and the special nature of the communication can be seen in the intentions of the speech acts (ibid.). As Austin (1980 [1962]) already noted, speech acts can be defined as expressions not used only to say things, but rather to do things à la I pronounce you husband and wife. Furthermore, such expressions cannot be assessed as true or false. Sager (1994: 24) points out that speech acts in LSPs do not make use of phatic or poetic communication, and often divide texts into different sections according to their
intentions, such as main body (informative), conclusion (evaluative) and recommendations (directive). By contrast, GLs can have mixed intentions in a single sentence. As far as style is concerned, LSPs tend to represent scientific-theoretic or informative styles whereas texts written in GLs usually represent informative or artistic style (Niemikorpi 1996: 106-107).

Schröder (1990: 9–10) emphasizes the fact that LSPs do not have anything to do with sociolects. The latter are sublanguages of a social or professional group meant to exclude people not belonging to the group whereas the aim of LSPs is only to make the communication in a given field as precise and effective as possible.

Another important feature of LSPs mentioned by Schröder (1990: 10) is the use of non-verbal communication. Diagrams, tables, formulae etc. are indeed good methods to make communication more effective, and Sager, Dungworth and McDonald (1980: 6, 15) consider their use to be a good demonstration of the systematic and functional nature of LSPs. Special languages can be considered international and so efficacy and systematic organization of information can be considered a prerequisite for successful communication. After all, LSPs are born out of a genuine need for one in specific fields.
2.4.1 The Language of Football as an LSP

The game of football was created in England several centuries ago, but it was only in the 1800s that it was turned from a village game into a real sport. In its early stages, the new sport was disparaged and ridiculed, and it had to compete with rugby. Despite the hardships, the English soon realized that football was a suitable game for their country’s rainy climate, and in the 1900s, the game had already become immensely popular. (Eskola et al. 2003: 17–18.) Nowadays, football is the world’s most popular game and, therefore, serious business. In other words, the game has turned from a hobby into a professional sport.

Being a professional sport, football has its own, distinctive rules, and this is partly why the game also has its own terms, such as free kick, offside, throw-in and corner kick. Naturally, quite many of these terms are used in other games as well. For instance, offside is used in ice hockey. Football terminology is also subject to change even though the changes are slow if compared to those in other LSPs.

However, a good example of a football term whose status has changed somewhat rapidly is libero which is Italian for free and refers to a defensive player who does not have a designated opponent. Liberos, or sweepers as they are also called, were still used quite often in the 1990s, but not many teams use one today, the main exception to the rule being some German teams. Even though the term per se has not lost its
meaning, you do not see it very often nowadays, and even when someone refers to a libero, they usually talk about a sweeper.

Non-verbal communication is quite common, for example, in rule books. The dimensions of the field of play, the goals, the ball etc. are often shown in figures and tables. Numbers as 4-4-2 and 4-5-1 may seem cryptic to a layman, but professionals know that they are referring to the number of defenders, midfielders and forwards on one team.

All three main styles, scientific-theoretical, informative and artistic, are used when writing about football. If LSPs are seen as representing scientific-theoretical and informative styles, rulebooks and different kinds of guides are sources wherein the special language of football should be seen in its purest form. However, the vast majority of football related texts consists of reports and articles coloured by the writers’ own views and estimates, as well as metaphorical language that quite often makes use of warlike imageries – which is probably typical of all kinds of sports discourse. In fact, this division between “pure” special language and texts bordering on GLs is the very principle according to which the research material analyzed in this study is categorized (see chapter 4).
3. PREVIOUS RESEARCH

Stylistic differences between LSPs and GLs have been studied many times before. Niemikorpi (1996: 43) finds that the number of nouns, adjectives and numerals in LSPs – 52.9% of all words – is 6.7% higher than that in GLs. By contrast, the number of finite verbs, pronouns, adverbs, and post- and prepositions in GLs – 34.5% of all words – is 5.8% higher than that in LSPs. He (Niemikorpi 1996: 45) also comes up with results which show that words, clauses and sentences tend to be longer in LSPs than in GLs.

Even though Niemikorpi’s results are based on Finnish texts, he (1996: 44–45) points out that similar results have been received from many other languages, such as English, Swedish, German and Russian, but they are not directly comparable to Finnish or each other because of the structural differences between the languages.

Sager, Dungworth and McDonald’s (1980: 234) findings on English seem to confirm Niemikorpi’s assumption. According to them, nouns and adjectives together form up to 60% of LSPs. Adverbs are much more common in GLs (8%) than in LSPs where their share is only about 4%. As a result of nominal style, verbs in special languages have much less communicative value than in GLs. In fact, there are two or three verbs in GLs for every verb in LSPs.
Gustafsson (1990) studies the role of lexical density as a style marker in LSPs. To be more specific, she studies five maritime documents written in English and the size and repeat rate of vocabulary in them as compared to GLs. Her results show that the repeat rate of vocabulary items is much higher in special languages than in GLs.

In her later study, Gustafsson (1993) determines that the large repeat rate in LSPs is due to the repetition of some lexical items. Interestingly enough, only 15% of all unique words in her LSP corpus are special terms, but they are also the words repeated most often. What is more, she concludes that the vast majority of the vocabulary could be considered general, and that the general words are more difficult to understand than the actual terms in her corpus. (Gustafsson 1993: 231.)

The languages of basketball and ice hockey have also been studied from the point of view of special languages. Korhonen (2000) studies the language of basketball, but instead of focusing on stylistic aspects, he emphasizes terminology. He finds out, however, that sentences in the language of basketball are not very long or complex, and that there is a large number of but-clauses, which can be explained by the evaluative and neutral tone of the analyzed text. More surprisingly perhaps, the number of interrogatives is relatively large compared to other LSPs – another feature reflecting the text’s evaluative nature.

Luokkala (1978) studies the special language of ice hockey in his master’s thesis. His corpus consists of both written and spoken texts. Even though Luokkala himself does
not call his study quantitative style analysis, a large proportion of the thesis focuses exactly on quantitative properties, such as the frequencies of nouns, adjectives, verbs, adverbs etc. The general conclusion in Luokkala’s work is that English ice hockey language can be distinguished from other varieties of English because it contains a large number of dynamic, descriptive and transitive verbs, and the frequency of both subordinate clauses and long sentences tends to be low. The main difference between the spoken and written forms of ice hockey language is that adjectives are not used very often in speech. (Luokkala 1978: 95.)

The present study resembles previous ones in the sense that its focus is mainly on quantitative properties and their effect on style. However, the present study also aims to show how these properties may affect readability, which makes it different from the studies introduced above. What is more, in the present study, style and readability are approached from a qualitative angle as well with the aid of a questionnaire, which is yet another feature not found in the previous studies.

When it comes to criteria set for word classes, this study and the previous ones are bound to differ from each other to some degree. Even though this makes it quite difficult to compare results, it is not altogether impossible, as different criteria are likely to affect the frequencies of some grammatical elements less than others. Nouns are a good example of these. Since there nonetheless are differences, the comparisons drawn in the present paper are mainly between different texts in the corpus.
4. MATERIAL AND METHOD

The material to be analyzed in this work consists of extracts from six different texts. I have divided these texts into three pairs, and I explain the basis for this below:

1) One Finnish and one English LSP text
2) One Finnish and one English text considered GL
3) One Finnish and one English text both of which can be seen as falling somewhere between LSPs and GLs

The first pair is formed by *Jalkapallosäännöt 2007*, the official rule book used by the Football Association of Finland, and *Laws of the Game 2006*, the official rule book by Fédération Internationale de Football Association (FIFA). As far as style is concerned, both the texts represent, first of all, scientific-theoretical and informative styles. Because the texts fall into the genre of rule books, it means that they are also formal and directive texts focused on the reader (see section 2.1.2). In other words, these texts are stylistically quite formal, and they should not contain any metaphorical language. Bearing in mind that both the texts have an official status, their target audience consists most likely of referees, officials, coaches, players and other professionals within the field. This is why these texts can be considered LSP texts.
Pair 2 includes extracts from the Finnish version of Wikipedia (http://fi.wikipedia.org, s.v. *jalkapallo* ‘football’), the free Internet encyclopedia that anyone can edit, and *Soccer for Dummies* (United States Soccer Federation, Inc. & Lewis, Michael 2000), a guide aimed at people who do not know much, if anything, about football. The style in these texts can be considered informative and, in the case of *Soccer for Dummies* at least, also artistic. The texts focus on the topic, so they are declarative and descriptive in nature. The texts can be seen as representing GL that is virtually free of excessive use of terminology. In other words, these sources could be used as the very first introductions to the game of football. Thus, these texts can be seen as representing GLs.

Finally, the third pair comprises a Finnish football magazine *Goal* 9/2006 and the magazine *Sporting Legends – World Cup 2006 Review*. Similarly to pair 2, these texts may be seen as representing artistic and informative styles, but the former is probably more emphasized in this pair than in pair 2. Resembling reviews to a large extent, the texts are also declarative and descriptive in nature – another feature in common with pair 2. Football magazines, however, are aimed at football fans, not beginners or laymen, which makes it possible to use relevant terminology. The language in these magazines is probably sports language at its most typical: rich in figurative language and different kinds of estimates. In other words, the writers have more stylistic freedom. The texts may also contain comments from players, coaches etc., and they are included in the analysis as well because they are an important part of the style of
the text. Besides, writers usually adapt comments to suit their own style. Hence, these texts fall into the third category.

The extracts to be analyzed were chosen to be used here because they were considered a good cross-section of the styles in each of the texts. All the extracts are about 2,000 words long, and they can be found in the Appendices. In the quantitative analysis, however, only running text is taken into account, so all headlines, captions, incomplete sentences and tables are ignored. They are present in the qualitative test because the readers read the original versions of each text. Even though only complete sentences are accepted, sentences do not have to begin with a capital letter nor end in a full stop, exclamation mark or a question mark. This means that even sentences beginning a list of, say, different kinds of criteria and ending with a colon are regarded as complete sentences if they make sense as solitary sentences. However, sentences in the middle of paragraphs cannot end in a colon, semicolon, hyphen, dash etc.

Contractions, or merged words, such as you’re, are counted as separate words. In other words you’re would be counted as two words. Abbreviations are considered single words, and are categorized according to their primary roles in the given contexts.

Of course, some extracts have been cut to meet the length requirement of roughly 2,000 words. It should be stressed that even though some parts are omitted in the
quantitative analysis, there are still about 2,000 words in each analyzed extract. In other words, each pair of extracts consists of about 4,000 words, which should be a large enough sample to reveal the differences within and between each pair.

This study differs from many other corpus-based studies in two essential ways. Firstly, unlike most present day corpora, the research material used in this study does not consist only of electronic texts. To be precise, three of the sources mentioned above, i.e. both rule books and the extract from Wikipedia, are in an electronic form while the other three are printed. Secondly, no statistical software or programs are used when the texts are analyzed. Instead, all data is gathered manually to ensure that each and every word falls into the right category in the quantitative analysis.

Stylistics is often based on comparison, and this study is no exception. As mentioned in chapter 3, the comparison will mainly be between different texts in the corpus, but also between the results found in the analysis and those found in previous studies. However, the previous results and the results found here are likely to be only partly comparable due to different criteria for word classes.

All the extracts are subjected to a two-part analysis. Firstly, they are studied with the aid of a style and readability test developed specifically for this thesis and based on widely accepted ideas about good readability (see section 2.1.3). This test focuses on quantitative stylistic features. In other words, in the analysis, style is a statistical phenomenon, not an impression on the reader (cf. Varantola 1984: 51). Secondly, a
group of readers tells its opinion about the texts by filling in a questionnaire also
developed specifically for this study. This should be a more modern means of testing
readability than different kinds of readability formulae introduced in section 2.1.3.
Naturally, the questionnaire is mainly qualitative in nature. Now it is time to take a
closer look at these methods.

4.1 Quantitative Test

The quantitative test draws on the ideas about good readability introduced in section
2.1.3. Hence, sentence length and the frequencies of nouns, verbs, adjectives,
pronouns and numerals are counted.

Sentence length is easy enough to understand – it is defined by the number of words
in each sentence. Words forming English compounds are considered a single word
only when they are not written separately. To be more specific, each and every word
is analyzed as an individual entity. Numerals refer both to numbers written as a word,
such as three, and to actual numbers. In this study, ordinals are also considered
numerals. Pronouns are seen as a single group, so they are not divided into any
subcategories. Other elements analyzed in the study require more explaining, and that
will be done next.

Adjectives are first counted as a single group, after which the frequencies of positive,
comparative and superlative forms are counted in proportion to the total number of
adjectives. Nouns are first counted all together, and then the frequencies of proper and common nouns are determined. Countable and mass nouns are not counted separately.

Verbs are first counted as a single group as well, and then the relation between the use of the passive and active voices is determined. However, no separation between finite and nonfinite, dynamic and stative, or full verbs, primary verbs and modal auxiliaries is made at any point since it has no relevance to the present study.

4.1.1 Problems with English Participles

In the English language, distinguishing between participles and adjectives can be quite difficult at times because they use the same suffixes –ing and –ed (and corresponding verb forms). Adjectives of this kind are therefore called participial adjectives. Recognizing participial adjectives is easy when the forms in question have no corresponding verbs, such as talented and *to talent.

In other cases, the presence of a direct object or a by-agent may help define whether the form in question is a participle or an adjective. In some cases, such as you are insulting me, it is easy to see that insulting is a participle. Naturally, the same applies to the house was painted red by John. The intensifier very, in turn, often indicates that the form in question is an adjective: The man was very offended. Should there be a by-agent, however, the form would be a participle, and in the absence of both an
intensifier and a by-agent, it is almost impossible to say which form is in question as in the man was offended. (Greenbaum & Quirk 1990: 133–134.) Should forms of the last kind be used in the analyzed corpus, the focus will be on the outcome of the process. In other words, offended would be seen as an adjective.

4.1.2 Problems with Finnish Participles

It may also be problematic to analyze participles in the Finnish language. There are two kinds of participles in Finnish:

1\textsuperscript{st} participle \ sanova  \ ‘saying’

Passive form \ sanottava  \ ‘to be said’

2\textsuperscript{nd} participle \ sanonut  \ ‘said’

Passive form \ sanottu  \ ‘said’

Problems arise because, as in English, participles may be used like adjectives. They can premodify nouns and function as complements as in \textit{puhuva poika} ‘speaking boy’ and \textit{laulamani laulut} ‘songs sung by me’. In the context of this study, all these forms are considered verbs excluding cases wherein the form has established its status as an adjective. The word \textit{viehättävä} ‘attractive’ is a good example of a participle that is regarded as an adjective.
4.1.3 Other Problems Related to Categorization

It is not possible to mention every possible problem that you may face when categorizing words into classes. This being the case, I can only refer to the sources consulted when facing a problem not mentioned above:

1) Finnish: Iso suomen kielioppi (Hakulinen et al. 2005)
2) English: A Student’s Grammar of the English Language (Greenbaum & Quirk 1990)

4.2 Qualitative Test (Questionnaire)

At the beginning of the questionnaire, the readers are asked to estimate their knowledge of football on a 3-point scale where the alternatives are poor, average and good. At the end of each text, there are 8 basic questions the readers are asked to answer without returning to the text. These questions do not change when the reader moves from one extract to another. The readers are asked to mark their answers on a 5-point scale where the alternative “I disagree” is at one end and “I agree” at the other. However, the scale is not used in conjunction with the first question, as in it, the readers are asked to mark the time spent on reading the text.

In addition to the basic questions, there are three text-specific questions about each text, and once again, the readers are asked not to refer back to the texts. These
questions test how well the readers have understood what they have read and how well they remember the text. These questions focus on the beginning, middle and end of each text, respectively. These questions are not judged by the 5-point scale, but the answers are simply considered correct or wrong. Each correct answer can be regarded as a sign of good readability and coherent style. At the end of the questionnaire, the readers are given a chance to freely comment on the text and its style.

The questionnaire is sent to altogether fifteen people with different backgrounds. Seven of these people are fluent in English, as they are either students of English Translation and Interpretation, students of English Philology, or have lived in an English-speaking country. Therefore, they should be suitable to read the texts written in English and answer the questions about them. The remaining eight readers read the Finnish texts. The people taking the questionnaire are either personal contacts or were found through them.

Because all respondents within the test group speak Finnish, the questionnaire is in Finnish. This original version, as well as its translation, are in the Appendices. However, for ease of reference, the English translations of the eight basic questions in the questionnaire are presented at the beginning of section 5.3 in conjunction with the actual analysis of the results of the questionnaire.
5. ANALYSIS

The next sections contain all the statistical data gathered from the quantitative analysis and the questionnaire.

5.1 Quantitative Analysis and Its Results

Let us discuss the data gathered from the quantitative analysis of each extract. The results can be seen in table 1 on the following page. There are many different points of analysis, and they are dealt with in their own sections. Average sentence length is the first point of analysis listed in table 1, so it is only appropriate to begin the discussion of the results with it. The word classes are then dealt with according to their frequency – in other words, from the largest to the lowest.
Table 1. Average sentence length, and frequencies of lexical items. 1 = Wikipedia, 2 = Säännöt 2007, 3 = Goal 9/2006, 4 = Soccer for Dummies, 5 = Laws of the Game, 6 = Sporting Legends.

<table>
<thead>
<tr>
<th>Extract</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of words</td>
<td>2030</td>
<td>1945</td>
<td>1941</td>
<td>2066</td>
<td>1973</td>
<td>1943</td>
</tr>
<tr>
<td>Points of analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average sentence length (words)</td>
<td>14.1</td>
<td>12.9</td>
<td>12.4</td>
<td>18.1</td>
<td>18.8</td>
<td>24.0</td>
</tr>
<tr>
<td>Adjectives (A) of all words</td>
<td>6.2%</td>
<td>5.1%</td>
<td>8.1%</td>
<td>5.8%</td>
<td>6.2%</td>
<td>4.0%</td>
</tr>
<tr>
<td>Positives of all A</td>
<td>77.8%</td>
<td>88%</td>
<td>87.9%</td>
<td>89.2%</td>
<td>96.7%</td>
<td>94.8%</td>
</tr>
<tr>
<td>Comparatives of all A</td>
<td>10.3%</td>
<td>7%</td>
<td>5.1%</td>
<td>5%</td>
<td>3.3%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Superlatives of all A</td>
<td>11.9%</td>
<td>5%</td>
<td>7%</td>
<td>5.8%</td>
<td>0%</td>
<td>3.9%</td>
</tr>
<tr>
<td>Nouns (N) of all words</td>
<td>46.4%</td>
<td>42.8%</td>
<td>40.5%</td>
<td>24.5%</td>
<td>29.4%</td>
<td>31.4%</td>
</tr>
<tr>
<td>Common nouns of all N</td>
<td>73.5%</td>
<td>96.5%</td>
<td>57.1%</td>
<td>89.2%</td>
<td>90.5%</td>
<td>57.1%</td>
</tr>
<tr>
<td>Proper nouns of all N</td>
<td>26.5%</td>
<td>3.5%</td>
<td>42.9%</td>
<td>10.8%</td>
<td>9.5%</td>
<td>42.9%</td>
</tr>
<tr>
<td>Verbs (V) of all words</td>
<td>18.5%</td>
<td>23.9%</td>
<td>21%</td>
<td>20.5%</td>
<td>19%</td>
<td>19.3%</td>
</tr>
<tr>
<td>Act. &amp; pass. voice ratio</td>
<td>2.8</td>
<td>5.6</td>
<td>8.5</td>
<td>9.5</td>
<td>1.5</td>
<td>7.6</td>
</tr>
<tr>
<td>Numerals (Nu) of all words</td>
<td>4.5%</td>
<td>2.2%</td>
<td>2.3%</td>
<td>1.7%</td>
<td>1.9%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Pronouns (P) of all words</td>
<td>6.4%</td>
<td>6.7%</td>
<td>6.6%</td>
<td>10.6%</td>
<td>6%</td>
<td>6.9%</td>
</tr>
<tr>
<td>A + N + V + Nu + P of all words</td>
<td>82%</td>
<td>80.7%</td>
<td>78.5%</td>
<td>63.1%</td>
<td>62.5%</td>
<td>64.8%</td>
</tr>
</tbody>
</table>
5.1.1 Average Sentence Length

As the reader can see in table 1 above, the average sentence length in the Finnish texts does not vary considerably. However, it is quite interesting that the average sentence length is not the longest in the extract taken from Säännöt 2007, which was regarded as an LSP text before the analysis, but in the extract from Wikipedia that, in turn, was considered GL. The average sentence in the Wikipedia extract is 14.1 words long, which is, in fact, quite a high figure for a text that is supposed to be GL. The average sentence in Finnish GLs is namely only 11.5 words long (Niemikorpi 1996: 45). Hence, it comes as no surprise that the average sentence in Säännöt 2007 is considerably shorter – only 12.9 words – than the average sentence in Finnish LSP texts that is 15.6 words long (ibid.).

The third Finnish extract, taken from Goal 9/2006, which was seen as falling somewhere between GLs and LSPs, has an average sentence length of 12.4 which is very close to a typical Finnish text written in GLs. Because the differences in average sentence length between the texts were not very significant, it is quite safe to say that sentence length cannot be used as a means to highlight stylistic differences between these texts. Thus, it is also evident that one cannot base his or her opinion on average sentence length when claiming that one of these texts is more readable than another.

When looking at the average sentence length in the texts written in English, one figure catches the eye immediately. The average sentence length in Sporting Legends
is significantly longer than in the two other texts. While the latter have a sentence length of about 18 words, Sporting Legends uses sentences that are 24 words long on average. This is quite close to the average sentence length of the press, which varies between just under 20 and a little over 21 words (see Francis & Kučera 1982: 552). Sporting Legends was seen as a text falling somewhere between GL and LSP texts, but when compared to its Finnish counterpart Goal, one notices something interesting. Compared to the Finnish football magazine, the average sentence is much longer in Sporting Legends. Whether or not this is a typical phenomenon between Finnish and English sports texts is almost impossible to say on the basis of merely two texts. However, one could assume that the typical sentence in Sporting Legends is so long that the text as a whole could be expected to prove less readable than the other two texts written in English. It is also a distinct stylistic marker.

5.1.2 Nouns

One of the most prominent features of the style employed in LSPs is the large frequency of nouns. Indeed, the Finnish extract labelled as LSP does contain a notable number of nouns, 42.8%. Bearing in mind once again that criteria may vary between different studies, Niemikorpi’s (1997: 43) statistics show that the percentage of nouns in Finnish LSPs is 39.3% on average. Hence, it is all the more surprising to see that the Finnish GL text contains even more nouns than the LSP text – over 46%.
Of the Finnish texts, the football magazine is the one with the smallest number of nouns, but it, too, reaches above 40%. Therefore, it is obvious that all the Finnish texts have a nominal style. Because there is a significant difference in the use of nouns between the Finnish texts, using the frequency of nouns to measure readability is quite well grounded. Thus, the lower the frequency of nouns is, the better the readability is.

The English texts, on the other hand, cannot be described as having a nominal style as the frequency of nouns varies only between 24.5% and 31.4%. In this case, it is the GL text, Soccer for Dummies, with the smallest number of nouns. However, this figure is quite low even for a GL text, especially if one is to believe the data gathered by Sager, Dungworth and McDonald (1980: 234) according to which the frequency of nouns in English GLs is 28% on average. On the other hand, the style of the rulebook, Laws of the Game, cannot be called very nominal, either, as nouns constitute only 29.4% of it. This is nowhere near the frequency, 44% that an English LSP should contain, at least according to Sager, Dungworth and McDonald (ibid.). The gap between 24.5% of nouns in Soccer for Dummies and 31.4% in Sporting Legends is large enough to be used as a basis when measuring readability.

When comparing the English texts to the Finnish texts, one notices that the English football magazine tends to make more use of nouns than the other English texts, while the situation with the Finnish texts is the exact opposite. A similar comparison
can be made with the Finnish and English GL texts. While the English one uses nouns quite sparsely, the Finnish one is very rich in them.

The results also show that the Finnish GL text makes much more use of proper nouns than its English counterpart: 26.5% to 10.8%. Another similar stylistic difference is seen between the LSP texts. 9.5% of the nouns in the English rulebook are formed by proper nouns, whereas in the Finnish text the figure is only 3.5%. Both the Finnish and English football magazines, on the other hand, are very rich in proper nouns, which may affect the readability of the texts if the reader is not familiar with these names.

5.1.3 Verbs

Due to the high frequency of nouns in the extract from Finnish Wikipedia, it is not surprising to find that the communicative value of verbs in the extract is relatively low if their frequency is used to measure it. Verbs constitute only 18.5% of the extract, whereas their frequency in the rulebook Säännöt 2007 is over 5% higher, 23.9%. The relation between the use of the active and passive voices is 2.8 in Wikipedia and 5.6 in Säännöt 2007, which is yet another peculiarity that does not follow the typical GL-LSP pattern. The football magazine Goal 9/2006, however, favours the use of the active voice even more often than Säännöt 2007, as the relation between actives and passives is as high as 8.5 while the percentage of verbs is 21%.
Again, it is the English texts that seem to adhere better to the GL-LSP pattern. Soccer for Dummies, which is seen as GL, contains more verbs, 20.5% than the LSP text, Laws of the Game, where their share is 19%. The active voice is used very often in Soccer for Dummies, 9.5 times more often than the passive. In Laws of the Game, the passive voice is used almost as often as the active, as the relation between them is only 1.5 in favour of the active voice. The football magazine, Sporting Legends, falls between the two because verbs constitute 19.3% of it while the relation between actives and passives is 7.6.

In both the English and Finnish texts, the differences in the relations between the use of the active and the passive voices are significant enough to be used when measuring readability. The frequency of verbs may also be used to show a difference between all the Finnish texts. When it comes to the English texts, however, it can only be used to show that the text labelled as GL may be more readable than the other two texts, which is due to the very small margin of .3% in the use of verbs between the latter two.

5.1.4 Pronouns

The frequency of pronouns remains almost identical when moving from one Finnish text to another. This frequency is about 6.5% in all the texts. The frequency of pronouns in the English rulebook and football magazine is about the same, 6% and 6.9%, respectively. Of course, the English texts contain more pronouns compared to
the Finnish ones, as the presence of articles and prepositions in the English texts affect the results. The text with the highest share of pronouns is Soccer for Dummies of which pronouns constitute 10.6%.

While it is difficult to say how pronouns affect readability, it is quite easy to ascertain that all the English texts use more anaphoric devices than the Finnish ones, and creating cohesion with the aid of pronominal anaphora is more typical of the style employed in the English texts.

5.1.5 Adjectives

LSPs should contain more adjectives than GLs (Niemikorpi 1997: 43). When contemplating the statistics in table 1, one sees that this seems to hold true as far as the English texts analyzed in this study are concerned. The percentages are 6.2% to 5.8% in favour of the LSP text, Laws of the Game. This difference is by no means significant considering the relatively small size of the corpus, and cannot be used to say which text is more readable or stylistically superior in this sense. What is significant though is the low frequency of adjectives in Sporting Legends, in which only 4% of all words are adjectives. This is quite an interesting detail as one would expect a sports magazine to contain much more adjectives. After all, 4% is a low figure for any kind of text – a factor which inevitably affects the style of the text and possibly even its readability.
When it comes to the Finnish texts, the situation is completely different. The text labelled as LSP, i.e. Säännöt 2007, has the lowest number of adjectives, 5.1%, while the highest number of adjectives is found in Goal, 8.1%, which is the counterpart of Sporting Legends. In other words, the Finnish football magazine seems to make much more use of adjectives than the one written in English. On the other hand, the relatively high frequency of nouns in the English football magazine may explain why the frequency of adjectives is as low as 4%. It is namely likely that the writer of the text has used adjective-like nouns, which has of course affected the results here.

As one can expect, the positive forms of adjectives are used most often in all the texts. What is especially interesting is that when it comes to the use of all different forms of adjectives, the Finnish and English texts rank in exactly the same way:

Positive forms:

<table>
<thead>
<tr>
<th>Finnish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>88% in Säännöt 2007 (LSP)</td>
<td>96.7% in Laws of the Game (LSP)</td>
</tr>
<tr>
<td>87.9% in Goal (LSP/GL)</td>
<td>94.8% in Sporting Legends (LSP/GL)</td>
</tr>
<tr>
<td>77.8% in Wikipedia (GL)</td>
<td>89.2% in Soccer for Dummies (GL)</td>
</tr>
</tbody>
</table>

Comparative forms:

<table>
<thead>
<tr>
<th>Finnish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.3% in Wikipedia (GL)</td>
<td>5% in Soccer for Dummies (GL)</td>
</tr>
<tr>
<td>7% in Säännöt 2007 (LSP)</td>
<td>3.3% in Laws of the Game (LSP)</td>
</tr>
</tbody>
</table>
5.1% in Goal (LSP/GL) 1.3% in Sporting Legends (LSP/GL)

Superlative forms:

Finnish                English
11.9% in Wikipedia (GL) 5.8% in Soccer for Dummies (GL)
7% in Goal (LSP/GL)     3.9% in Sporting Legends (LSP/GL)
5% in Säännöt 2007 (LSP) 0% in Laws of the Game (LSP)

Hence, it seems that both English and Finnish football texts written in GLs seem to favour the use of positives and superlatives more than texts labelled as LSPs. It must be said, however, that with the exception of Sporting Legends, none of the texts analyzed here can be said to employ adjectival or non-adjectival style. One cannot rely on previous studies here, either, because the criteria used in them are most likely to differ from the ones used here. Hence, suffice it to say that compared to the Finnish texts, the English ones tend to make more use of adjectives, with the exception of the football magazine, of course.

5.1.6 Numerals

Of the Finnish texts, numerals are used most often in the extract from Wikipedia, where their share is 4.5%. This is a little surprising considering that the text is supposed to be GL. Indeed, its English counterpart, Soccer for Dummies, has the lowest frequency of numerals of the English texts, only 1.7%. Another text that has a
relatively high number of numerals in it is the English football magazine Sporting Legends with 3.2%.

Since a high number of numerals in a text may make reading and remembering more difficult, it is worth taking their frequency into account when measuring readability. In this respect, the difference should be quite evident between the Finnish GL text and the other two Finnish texts, as well as between the English football magazine and the other two texts. As a general conclusion, it might be noted that the frequency of numerals is likely to depend on both the topic and the field. It is probable that the frequency of numerals is somewhat lower in a book dealing with discourse studies, for instance.

Before moving on to the qualitative analysis and its results, it is in order to sum up the results of the quantitative analysis and see how readable the texts can be expected to be when compared to one another.

5.2 Expectancy in Readability

The easiest way to sum up the results regarding the readability of all the texts is to rank them according to all the elements that were seen as having an effect on readability in the sections above. Hence, the following ranking of the readability of the Finnish texts is based on the frequencies of adjectives, nouns, verbs and numerals, as well as the relations between the use of the passive and active voices.
2. Säännöt 2007 (LSP)
3. Wikipedia (GL)

The following ranking of the English texts is in turn based on average sentence length, the frequencies of adjectives, nouns, verbs and numerals, as well as the relations between the use of the passive and active voices.

1. Soccer for Dummies (GL)
2. Laws of the Game 2006 (LSP)
3. Sporting Legends (LSP/GL)

It should be mentioned that the differences were very significant in both groups. This is quite interesting as the rankings are very different from each other. In fact, the order of the texts is reversed when one moves from the ranking of the Finnish texts to that of the English. However, these rankings are anything but conclusive, but it is nevertheless interesting to see whether they correlate with the results received from the questionnaire.
5.3 Qualitative Analysis and Its Results

In order to understand the data gathered from the questionnaire, it is necessary to know what exactly was asked in it. The translations of the eight basic questions can be seen below. The translations of the text-specific questions, i.e. questions 8-10, are available in the Appendices.

The time spent on reading (by the accuracy of one minute):

1. The text was understandable and easy to read.
2. The text was interesting.
3. The writer’s style was pleasant.
4. There were words or phrases in the text that I did not understand.
5. There were sentences or sentence structures that were hard to understand.
6. The text was consistent.
7. The text was divided well into paragraphs and different subjects.

The questionnaire was sent to altogether 15 people, 8 of whom were asked to answer questions regarding the Finnish texts, and 7 to answer questions about the texts written in English. To be exact, 5 people from both groups turned the questionnaire in, which means that 62.5% of the “Finnish” group and 71.4% of the “English” group returned the questionnaire.
The average answers to the eight basic questions asked after each text can be seen in table 2 below. An average of 1 translates into “I disagree”, 2 into “I disagree in part”, 3 into “I do not know, 4 into “I agree in part” and 5 into “I agree”. The question about the time spent on reading is, of course, an exception.

Table 2. Average answers to the basic questions asked in the questionnaire. 1 = Wikipedia, 2 = Säännöt 2007, 3 = Goal 9/2006, 4 = Soccer for Dummies, 5 = Laws of the Game, 6 = Sporting Legends.

<table>
<thead>
<tr>
<th>Extract</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. time spent on reading (min.)</td>
<td>20.4</td>
<td>20</td>
<td>15.6</td>
<td>13.8</td>
<td>15.8</td>
<td>18.8</td>
</tr>
<tr>
<td>Question no.</td>
<td>avg.</td>
<td>avg.</td>
<td>avg.</td>
<td>avg.</td>
<td>avg.</td>
<td>avg.</td>
</tr>
<tr>
<td>1</td>
<td>3.6</td>
<td>3.8</td>
<td>3.4</td>
<td>5</td>
<td>4.8</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2.8</td>
<td>3.4</td>
<td>4.4</td>
<td>4.4</td>
<td>4.2</td>
</tr>
<tr>
<td>3</td>
<td>3.6</td>
<td>4</td>
<td>3.8</td>
<td>4.8</td>
<td>4.4</td>
<td>3.8</td>
</tr>
<tr>
<td>4</td>
<td>1.8</td>
<td>3</td>
<td>2.2</td>
<td>3.4</td>
<td>2</td>
<td>2.4</td>
</tr>
<tr>
<td>5</td>
<td>2.8</td>
<td>2</td>
<td>2</td>
<td>1.2</td>
<td>1.2</td>
<td>2.4</td>
</tr>
<tr>
<td>6</td>
<td>4.6</td>
<td>4.4</td>
<td>4</td>
<td>5</td>
<td>4.8</td>
<td>4.4</td>
</tr>
<tr>
<td>7</td>
<td>4.4</td>
<td>4.4</td>
<td>3.8</td>
<td>5</td>
<td>4.8</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Finally, the results gathered from the last three questions, i.e. questions 8-10, can be seen in table 3 below. These questions dealt with the content of the texts, and the answers were judged simply as correct or wrong.

<table>
<thead>
<tr>
<th>Extract</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question no.</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td>60</td>
<td>40</td>
<td>100</td>
<td>100</td>
<td>60</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>100</td>
<td>40</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
<td>80</td>
<td>60</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

5.4 Discussion of the Results of the Qualitative Analysis

The first question the readers were asked to answer was about the time spent on reading each text. When looking at the results of this question, one notices that, in the Finnish texts, the reading speed seems to increase when moving from the GL text to the LSP text and finally to the text seen as falling between GLs and LSPs. In the English texts, the order is reversed, and is thus more natural, as GL texts should be easier to read than LSP texts. Having said that, it should be stressed that the average times spent on reading the Finnish GL text and LSP text were virtually the same.

Indeed, according to the results of the questionnaire, the English GL text seems to be somewhat easier to read and understand than its Finnish counterpart. All that read the English text answered “I agree” to the first question, “The text was understandable
and easy to read”, whereas the average answer of the readers of the Finnish texts situates between “I do not know” and “I agree in part.” A similar difference is seen when contemplating the average answers to question three, “The writer’s style was pleasant”.

According to the readers, the English GL text contains more hard-to-understand words than the Finnish GL text, while the latter was seen as containing more cumbersome phrases and sentence structures. The first phenomenon can be explained, at least partly, by the fact that some of the words in question are not part of the readers’ English vocabulary while they might recognize the term itself if it was in Finnish. In fact, two readers suspected that this was indeed the case when given a chance to comment freely on the text. Hence, one should not place too much emphasis on this difference between the GL texts. When it comes to the text-specific questions, there is no difference between the texts, in that only one reader answered one of the questions wrong in both groups.

The differences between the LSP texts were equally clear, and again, in favour of the English one. With the average answer amounting roughly to “I agree”, the English text was easier to read and understand than the Finnish one. The readers did not consider the latter to be very interesting, either, as the average answer to question 3, “The text was interesting”, was only 2.8, “I do not know”. This cannot be said about the English text, as its average was as high as 4.4, which places it between “I agree in part” and “I agree”. 
The Finnish LSP text was also seen as being somewhat harder to read as far as the vocabulary, phrases and sentence structures were concerned. This is supported by the fact that none of the readers of the English text answered any of the text-specific questions wrong, while only two readers succeeded in answering correctly to all the questions about the Finnish text.

The results received from the football magazines match each other quite well, but there are a few significant differences. The English text was yet again considered better as far as style and understandability were concerned, but also in terms of how well the text was divided into paragraphs and different subjects. The results also show that the text-specific questions asked about these texts were by far the hardest in the questionnaire. The only question that all readers managed to answer correctly was the last question asked about the English text. Before comparing the results received with the two methods to each other, it is in order to find out how previous knowledge affected the readers’ reading experience.

5.4.1 The Effect of Previous Knowledge on Readability

The effects of previous knowledge of football were very clear in the analysis. The better a reader estimated their knowledge to be, the better they rated each text and were able to answer text-specific questions correctly.
What is surprising though is that previous knowledge did not seem to result in high reading speeds. On the contrary, in fact – those who reported to have a good knowledge of football spent much more time on reading the texts than those with a poor or average knowledge. Some readers commented that they had found the extracts from the football magazines very dull and claimed that the contents of the texts were very hard to remember because of the high frequency of proper nouns in them. Yet these readers read these texts very quickly. This can only lead to the conclusion that reading speed is not a reliable indicator of good readability. This is also supported by the fact that the readers who read the extracts from the football magazines the fastest, did not answer the content-related questions correctly as often as the slower readers.

5.5 Comparison between the Methods

To compare the results provided by the two completely different methods, it is necessary to rank all the texts on the basis of the results provided by the questionnaire. The rankings are then compared with the ones presented in section 5.2.

As mentioned in section 5.4.1, reading speed proves to be an unreliable indicator of readability in this study, so it is not taken into account in the ranking. Questions 1-10, on the other hand, can be used for this purpose. If the difference in average answers
to questions 1-7 is only between .0 and .4 points, it is considered too insignificant to make a difference between the texts. The rankings can be seen below.

Finnish texts:
1. Wikipedia (GL)
2. Säännöt 2007 (LSP)

English texts:
1. Laws of the Game 2006 (LSP)
2. Soccer for Dummies (GL)
3. Sporting Legends (LSP/GL)

It seems that the readers disagree quite strongly with the results of the quantitative method. This is most evident when comparing the rankings of the Finnish texts, as their order is reversed when moving from the ranking based on the quantitative analysis to the qualitative ranking. The only text that has the same position in both rankings is Sporting Legends, i.e. the English football magazine.

Even though the quantitative approach revealed some interesting, seemingly covert qualities that should have affected style and readability in one way or another, they did not seem to have a significant effect on the readers. For example, the readers of the Finnish texts seemed to like the GL text the most, and they also remembered its
content most accurately, even though the quantitative test showed that almost every other word in the text was a noun. The readers of the English texts, on the other hand, favoured the LSP text and remembered its content perfectly despite the fact that the passive voice is used almost as often as the active in the text, and the frequency of nouns is relatively high. However, it would appear that the high frequency of proper nouns in both football magazines seemed to have a great effect on those readers with a poor or average knowledge of football. These readers found the magazines less readable than the other texts – as was suspected in the quantitative analysis.

The fact that all the subjects were Finnish speakers has, without a doubt, had some kind of effect on the results of this study. However, using only Finnish speakers was not unwarranted. English can be seen as modern day Latin, or lingua franca, that belongs to everyone, not only to those who happen to speak it as their mother tongue. Further, non-natives form the majority of the readers of the English rule book Laws of the Game 2006 and the magazine Sporting Legends. In addition, the third English primary source used in the present study is a guidebook for football laymen written in very simple language. After all, all the subjects that read the English texts are credible speakers of English with almost native-like to native-like skills. Having said that, I do regret that it was not possible to use native speakers in this study alongside the Finnish ones, as it surely would have offered valuable insights into the matters at hand.
Finally, it is time to conclude by summing up the findings and seeing if the approaches applied here offered any answers to the research questions listed at the beginning of this study.
6. CONCLUSION

Regarding the research questions introduced in chapter 1, it is probably wisest to begin by answering the third question, as it also helps to answer the other questions. The results of the quantitative and the qualitative analyses do not support each other well. This being the case, emphasis must be placed on the latter, as style and readability are based on interaction between the text and the reader.

As already mentioned in section 2.4, it may be argued that there are no special languages, only special vocabularies. According to the respondents of the questionnaire, the football magazines were difficult to read because of the large number of proper nouns in them. This made the readers also claim that the texts were uninteresting and their content hard to remember. This was indeed the case with the Finnish magazine as well even though it ranked first in style and readability in the quantitative analysis. This naturally leads to the conclusion that if there was a special language of football, proper nouns would seem to be to it what terms are to other LSPs. That is to say that the mere quantity of proper nouns may make the reading of football magazines seem as difficult as that of more traditional LSP texts.

Furthermore, since these magazines proved much harder to read and understand than the rulebooks, it might be said that the magazines represent the special language of football better than the rulebooks even though the former may seem very GL-like at
times. The readers with a good existing knowledge of football, however, considered the magazines only slightly less readable than the other two texts on average, and they did not make many mistakes in the text-specific questions, either. Yet again, it must be noted that a native speaker familiar with the conventions of the language might have seen the texts in a different way.

In the quantitative analysis, the English texts seemed to follow the GL-LSP pattern somewhat better than the Finnish ones. The Finnish GL text in fact seemed less GL-like than the other two texts, but the readers claimed its style and readability to be much better than those of the other two. On the other hand, the readers thought the style and readability of the English LSP text to be the best.

Having arrived at such contradictory results, I can but say that the differences between different varieties of the language of football remain quite unclear. However, there are a few details that should be mentioned. It seems that the language of football as a whole does not make use of very long or complex sentences – this is the case, at least, when LSPs are used as a point of comparison. The only exception to this is the English football magazine, as evidenced by both the quantitative and qualitative test. In other words, the language of football seems to compare in this sense quite well with the language of other sports, such as ice hockey and basketball.

The English magazine also has another feature that makes it different from the other texts, and this may interest translators. This feature is the very low frequency of
adjectives, and it was suggested in section 5.1.5 that this might be due to the use of adjective-like nouns. Since this may affect readability, it might be wise to use real adjectives instead. By comparison, all the English texts contain more adjectives than the Finnish ones, so translators would do wisely to preserve them in their translations, especially when one bears in mind that Finnish football texts seem to have a very nominal style. Finally, translators should pay attention to the use of anaphoric devices, as English texts seem to make much more use of them than Finnish ones. Instead of repeating a noun time and again, one might replace it with a pronoun to create cohesion and to get rid of unnecessary nouns.

If one was to study the language of football further in a similar fashion, they might want to see rulebooks and magazines as a single group and compare it to texts written in GLs. Furthermore, if a quantitative approach should be used, it should be more refined than the one used here, and the corpus should also be somewhat larger, so as to reveal more tangible features. Having said that, this kind of approach should be combined with a qualitative one, as it is only reading that shows the true nature of a text. Reading speed, however, may not be a reliable indicator of good readability, as was discussed in section 5.4.1. Finally, the test readers should preferably be native speakers of the language the texts have been written in.
REFERENCES

Primary Sources


*Sporting Legends. World Cup 2006 Review.* Bath: Anthem.


Secondary Sources


