



LUND
UNIVERSITY

Naming A World of Things

An investigation of the development of noun phrases in written expository discourse, from adolescents to expert writers

Pia Gustafson

Supervisor: Victoria Johansson

Centre for Languages and Literature, Lund University

MA in Language and Linguistics, General Linguistics

SPVR01 Language & Linguistics: Degree Project – Master's (Two Years) Thesis, 30 credits

September 2016

Abstract

One component in writing development is a more efficient use of syntactic tools. Later syntactic development, specifically that of the noun phrase, is an understudied area, notably so for Swedish. The demand for specified and concise information in an expository text, and the availability of more planning time in the written modality, are important factors contributing to more complex noun phrases. The aim of this study is to analyse noun phrase length, lexicality and complexity in age-related and expert development.

This corpus-based study consisted of 96 expository texts written by groups of writers 10, 13, 15 and 17 years old, by adult university students and by adult expert students. All NPs (N=8670) – at least one pronoun or lexical noun plus modifiers – were analysed as lexical or pronominal, simple or complex, and NP length was calculated. Finally, development was analysed over syntactic constituent (Subject, Object or Other).

Results show that NPs are significantly less pronominal, longer, and more complex in older age groups, and most saliently in the subject constituent. However, two important implications are that age-related development is not straightforwardly linear, and that expert development results in *less* complexity. Results are discussed with respect to the concept of writing expertise, written modality characteristics and general cognitive development.

Keywords: later syntactic development, writing expertise, noun phrase lexicality, noun phrase length, noun phrase complexity

Acknowledgements

My first and heart-felt thank you goes to my supervisor, Victoria Johansson. Thank you, Victoria, for being so generous – generous with your time, your knowledge and experience, your material, your contacts, your demands as well as your sense of humour and your encouragement. Thank you, too, Joost van de Weijer, for expert guidance in statistics, and not least for helping me out with calculating noun phrase length. And thank you, all of you: my other teachers, colleagues and friends at the Linguistics and Swedish department at Lund University. Katarina Lundin deserves a special mention for valuable comments on the text, and Arthur Holmer for a discussion on typological differences in noun phrases. I've learnt a lot from you all. And writing this thesis has been such an enjoyable way to pass the time.

Table of contents

Tables	vi
Figures	vii
1 Introduction	1
1.1 Aim and hypotheses	4
1.2 Definitions and limitations	6
2 Background	8
2.1 Patterns in later syntactic language development	8
2.1.1 Increased length of the T-unit and clause	9
2.1.2 Increased use of subordination	9
2.1.3 Critique of the subordination and length indices	10
2.1.4 A note on lexical development and the noun phrase	10
2.2 The noun phrase in developmental research	11
2.3 Subject noun phrases in developmental and cognitive research	14
2.4 The development of writing expertise	17
2.5 The written modality and the noun phrase	18
2.6 The expository text type and the noun phrase	19
3 Method	21
3.1 Corpora	21
3.2 Participants, ethical issues and procedure in the original studies	21
3.3 Overview of participants and data for the present analysis	22
3.3.1 Participants	22
3.3.2 Material	23
3.4 Coding	26
3.4.1 Lexical and pronominal noun phrases	26
3.4.2 Noun phrase length.....	27
3.4.3 Simple and complex noun phrases.....	27
3.4.4 Clause constituent	28
4 Results	30
4.1 Lexicality – pronouns and lexical nouns	30
4.1.1 Lexicality overall – quantitative analysis	30
4.1.2 Lexicality across clause constituent – quantitative analysis.....	31
4.1.3 Subject noun phrase lexicality – quantitative analysis.....	33
4.1.4 Object noun phrase lexicality – quantitative analysis.....	33
4.1.5 Other noun phrase lexicality – quantitative analysis.....	34
4.2 Discussion of noun phrase lexicality results	35
4.3 Noun phrase expansion – length in number of words	37
4.3.1 Noun phrase length – overall quantitative analysis	37
4.3.2 Noun phrase length across clause constituent – quantitative analysis.....	38
4.4 Discussion of noun phrase length results	39
4.5 Simple or complex lexical noun phrases – complexity	42
4.5.1 Noun phrase complexity – quantitative analysis	42
4.5.2 Complexity across clause constituents – quantitative analysis.....	43

4.5.3 Discussion of noun phrase complexity results.....	45
5 Concluding discussion	47
5.1 Age-related development and the noun phrase.....	47
5.1.1 Timing.....	47
5.1.2 The complexity aspect.....	48
5.1.3 Development is not straightforward – the case of the 13-year-olds.....	49
5.1.4 University students best in class.....	50
5.2 Expert writing and the noun phrase	51
5.3 Method discussion.....	52
5.4 Concluding remarks, further questions and directions for further research.....	53
References	55

Tables

Table 1. Noun phrase categories.	6
Table 2. Overview of participants.	23
Table 3. Overview of material.	24
Table 4. Overview of noun phrases.	25
Table 5. Coding of noun phrases.	27
Table 6. Lexicality: percentage lexical and pronominal noun phrases.	30
Table 7. Lexicality: percentage lexical noun phrases across clause constituent and group.	32
Table 8. Length of lexical NPs.	38
Table 9. Percentage complex noun phrases across constituent and group.	42

Figures

Figure 1. Text by a 10-year-old.....	2
Figure 2. Text extract by an adult university student.	3
Figure 3. Text length in number of words.....	25
Figure 4. Lexicality: percentage lexical and pronominal NPs overall.	31
Figure 5. Lexicality: lexical and pronominal noun phrases across clause constituent.....	32
Figure 6. Lexicality: lexical and pronominal noun phrases as Subjects.....	33
Figure 7. Lexicality: lexical and pronominal noun phrases as Objects.	34
Figure 8. Lexicality: lexical and pronominal noun phrases as Other constituents.....	34
Figure 9. Noun phrase length in number of words.....	38
Figure 10. Complexity: complex and simple noun phrases overall.	43
Figure 11. Complexity: complex and simple noun phrases as Subjects.....	43
Figure 12. Complexity: complex and simple noun phrases as Objects.	44
Figure 13. Complexity: complex and simple noun phrases as Other constituents.....	44

1 Introduction

The noun phrase is key to the linguistic transfer of information in Swedish, and saliently so in text types like the written expository. The topic of the present master thesis is an investigation of the later development of noun phrase usage, addressing age-related as well as expertise-related evolution. Several scholars (Hunt 1965, Perera 1984, Scott 1988, Myhill 2009) highlight the fact that the area of later syntactic development is under-researched. Additionally, they point to phrase level constructions and frequently to the noun phrase as potentially informative structures for analysis.

In my BA thesis I investigated pause patterns during the writing process (Gustafson 2012). In that study, pausing time spent in different syntactic contexts was compared for three age groups – 13-year-olds, 15-year-olds and 17-year-olds. The results indicated that 17-year-olds made a greater cognitive effort (demonstrated as significantly longer pausing time during writing) in a phrase context, as distinct from that of the paragraph, clause or word, and compared to the younger age groups.

Shortly thereafter, we started exploring the characteristics of writing expertise in the Expert Writing project (Johansson, V. Expert writing – divine inspiration or hard work? Swedish Research Council: VR2011–2290, in progress). One of the findings was that adult expert writers used less complex syntax than an adult control group (Gustafson, Johansson, Johansson, Wengelin & Frid 2014). Ronald Kellogg (2008) emphasizes reader-awareness as an essential component of expert writing. Our interpretation of the results was that less complex syntax could well be one effect of that greater reader-awareness.

These aspects made me curious and convinced me that an analysis of noun phrases produced by different age and expertise groups might yield interesting insights into later syntactic development in written language. For instance, I asked myself whether 17-year-olds produce phrases that are substantially more complex than those of younger writers, and whether expert writers produce other kinds of phrases than typical adults. As a consequence, the present study explores the noun phrases in expository texts composed by individuals from six different age and expertise groups: from 10-year-olds to adult university students, completed with a group of students enrolled in a programme to become professional writers.

We might ask ourselves why we would want to investigate later syntactic development to begin with. The first answer, as hinted at above, is that we really do not know a lot about it – the area cries out for basic investigation, and not least in Swedish. Most previous research has been carried out in English, as becomes clear below. An important assumption underlying the current study is that the time from young adolescence and throughout adulthood entails a substantial syntactic growth, most worthy of attention, though not as conspicuous as early development. Not until the subtleties of written language and later-learned text types are scrutinized does this evolution become obvious. Cheryl Scott asserts that later language development simply cannot be measured by “the presence or absence of high-frequency

structures, but with the gradual acquisition of low-frequency structures and the ability to form unique combinations of structures” (Scott 1988:50). She adds that the competence to adjust syntax to new contexts, as well as the ability to use previously acquired structures for novel functions, are crucial components of that development. Employing an expanded, complex noun phrase rather than a simple one, or choosing a lexical noun phrase instead of a pronominal one, especially as subject constituent, can be examples of those less frequent, more specific or new structures that Scott refers to. The second answer to the why-question concerns the application of knowledge gained through basic research; precise knowledge about linguistic development is a prerequisite for challenging and supporting the language development of students at every stage.

Noun phrases do not only constitute the majority of phrases, but words within noun phrases are a substantial part of the total number of words, and most certainly in informative texts. To gain an understanding of this state of things we take a preliminary look at two text samples. These were produced by a 10-year-old (Figure 1) and an adult university student (Figure 2) respectively, and indeed give the immediate impression of distinct noun phrase patterns, in several respects.

Problem hade **dom två pojkarna som bråkade**. **Det** kunde berott på att **dom** gillade **samma tjej**.
Ibland kan **bästa kompisar** splittras på **ett sånt här bråk**.
Hur **man** ska lösa **det** kan inte **jag** förstå.
Jag har själv haft **ett sånt problem**. **Svartsjuka** kan leda till **bråk**. **Man** blir arg och irriterad och **man** kan vara väldigt sur på **många** när **ett sånt problem** uppstår. Eller när **man** går förbi **varandra** och så dunkar **axlar** mot **axlar**. Vad fan gör **du**? Se **dig** för. **Jag** ska se **mig** för **det** är ju **du** i så fall.
När **ett sånt problem** uppstår försöker **man** klart hjälpa till.
Men ibland kan **man** få **en smäll**. **Man** hämtar då **en rastvakt**. (wgo7fCEW)

Figure 1. Text by a 10-year-old, noun phrases are marked in orange. Spelling mistakes in the original text were corrected.¹

To start with Figure 1, which represents the complete text by a 10-year-old, more noun phrases are made up of single pronouns than of lexical nouns. Instances of pronouns are *de* (‘they’), and *dom* (‘them’) in the second sentence and *man* (generic pronoun ‘one’), *det* (‘it’),

¹ A problem had the two boys who were having a row. It could be because they liked the same girl. Sometimes best friends can be split up by such a row. How one is to solve it I can’t understand. I have had such a problem myself. Jealousy can result in a row. One becomes angry and irritated and one can be very cross with many people when such a problem occurs. Or when you pass each Other and then shoulders collide with shoulders. What the hell are you doing? Watch out. I am to watch out then it’s you. When such a problem arises obviously one tries to help. But sometimes one gets a wallop. One then finds a school break monitor.

and *jag* ('I'), in sentence number four. Secondly, the lexical nouns that are produced are predominantly short and frequent, and the same key lexemes occur repeatedly, e.g. *bråk* ('fight'), *tjej* ('girl') and *problem/ett sånt problem* ('problem/such a problem'). Thirdly, the lexical nouns are simple – only one of them includes a post-positioned modifier, in this instance a relative clause – *dom två pojkarna som bråkade* ('the two boys who were fighting'). Finally, we recognize a pattern where, frequently, the realization of the subject function tends to be a simple pronoun, whereas the lexical noun phrases that occur appear to favour other clause constituents (adverbial: *ett sånt här bråk* 'a fight like this', object: *en rastvakt* 'an adult school break monitor').

Mobbning är ett tema som med jämna mellanrum lyfts upp i media och som således då och då blir föremål för en samhällelig diskussion. Tyvärr möter de flesta av oss fenomenet betydligt oftare, i vår vardag. Tidigare tänkte man främst på barn och ungdomar då man hörde ordet. Idag har dock en annan kategori av mobbning blivit uppmärksammas, nämligen vuxenmobbningen. Precis som barn och ungdomar, kan vuxna frysa ut varandra och tillämpa olika metoder av fysisk eller psykisk terror för att skapa sociala skillnader. Vuxenmobbningen kan därför egentligen inte räknas som en från barn- och ungdomsmobbningen skild kategori av fenomenet mobbning. Det är fråga om samma drivkrafter och i grunden även samma metoder. Skillnaden ligger bara i människornas ålder samt därmed förbundna maktmedel.
[wu15mCEW]

Figure 2. Text extract by adult university student, noun phrases are marked in orange.²

The second example (Figure 2) is an extract from a text written by an adult university student under experimental conditions similar to the text by the 10-year-old above (see the Method section). What we see, in marked contrast to the preceding text, is a number of instances of lexical, expanded and complex noun phrases constituting various clause constituents. One instance is the subject of the fourth sentence *en annan kategori av mobbning* ('another category of bullying'). We also notice that noun phrases occupy a greater proportion of the text extract as a whole.

² Bullying is a theme that at regular intervals are highlighted in the media and which in this way now and then becomes the topic of a social discussion. Unfortunately most of us encounter the phenomenon considerably more often, in our everyday life. Previously one thought mostly of children and adolescents when one heard this word. Today another category of bullying gets attention, namely adult bullying. Just like children and adolescents, adults may ostracize each Other and apply different methods of physical or psychological terror to create social distinctions. Adult bullying therefor cannot really be counted as a category of the phenomenon of bullying distinct from the one of children and teenagers. It is a question of the same motivations and basically also the same methods. The difference is just in the age of the people and there to related methods of power.

As stated, the current study investigates on the one hand, age-related noun phrase development in writing, from 10 years of age and up to adulthood. However, two distinct groups of adults with different goals for their writing are also compared: “typical students”, and students participating in a creative writing program, educating themselves with the explicit aim of acquiring expert writing skills. In the present work, this development is termed expert development, or the development of expertise. Kellogg Hunt (1970) maintains that syntax develops if it becomes part of the practice and profession of the individual. Ronald Kellogg (2008), without addressing syntax specifically, maintains that only those who write regularly as part of their profession, or those who practise extensively to become professional writers, will produce expert writing. He compares it to how expertise is gained in other fields, like playing a musical instrument. Consequently, throughout the thesis, we discuss the two types of development – one that applies to “typical” adolescents and students in a literate society, taking place with increasing age, on the one hand, and one expert development, applying merely to those training the craft as such, and with writing skills as their specific goal, on the other.

The present investigation is carried out within the field of psycholinguistics, and more specifically the inquiry into cognitive processes underlying the development of writing and writing expertise. Syntactic development has often been described within a generative framework, but in this work, the theoretical account consists instead of cognitive–functional models and theories. A bottom-up, data-driven approach is used, taking as its starting point the actual production of noun phrases in written, experimentally elicited expository discourse data, exemplified in Figures 1 and 2 above.

1.1 Aim and hypotheses

The aim of this study, accordingly, is to investigate developmental changes in the noun phrase (NP, as defined in section 1.2), correlating with the writer’s increased age, on the one hand, and with expertise, on the other. Following previous research (summarized in the Background section), and the impression based on samples like the ones in Figures 1 and 2, changes are analysed from the aspects of lexicality, length and complexity. Furthermore, changes in those aspects are analysed across clause constituent (Subject, Object or Other). Since previous research is based on English, albeit typologically closely related to Swedish, the study is explorative to a certain extent. However, five predictions are made, and the rationale for each one is given below. To start with, four hypotheses state that with age,

1. noun phrase lexicality – the proportion of lexical noun phrases – increases, as the proportion of pronominal noun phrases decreases,
2. noun phrase length – measured in number of words – increases,
3. noun phrase complexity – the proportion of complex lexical noun phrases – increases, as that of simple noun phrases decreases,

and,

4. those developments of lexicality, length and complexity are distinct depending on the syntactic function of the noun phrase (Subject, Object or Other)– clause constituent.

The fifth hypothesis concerns itself with expert development, and predicts that with the development of writing expertise,

5. the length and complexity of noun phrases decrease.

The first hypothesis, on the development of lexicality, was generated on the basis of a number of studies that conclude that to an increasing extent, noun phrases produced by older writers generally, in written language and certain contexts specifically, are constituted by lexical nouns rather than by pronouns (e.g. Hunt, 1965; Scott, 1988). For examples of this salient distinction, study once more Figures 1 and 2. The numerous instances of single pronouns in all clause constituents (*de* ('they'), *dom* ('them'), *man* (generic pronoun 'one') et cetera) in the text written by the 10-year-old (Figure 1) contrast sharply to the number of specific lexical noun phrases in the adult text sample (*mobbning* ('bullying'), *vuxenmobbning* ('adult bullying'), *fenomenet mobbning* ('the bullying phenomenon')). Consequently, the first hypothesis is corroborated if, with age, the percentage noun phrases with a lexical head increases significantly at the expense of that of noun phrases with a pronominal head.

The second hypothesis concerns the length of lexical noun phrases in number of words, and predicts an expansion with age. The analysis of noun phrase length, thus, does not include noun phrases with a pronominal head. In previous research (e.g. Ravid & Berman, 2010) noun phrase expansion is viewed as the effect of a growing capability of packaging more information within each phrase by employing pre- and post-positioned modifiers. Noun phrase length, therefore, can be viewed as a complementary measure of complexity. The second hypothesis, logically, is confirmed if the mean length of all lexical noun phrases produced by the younger age groups is significantly shorter than the mean length of those produced by adults.

Thirdly, the prediction is made that the complexity of lexical noun phrases increases. As with the length measure, complexity does not include noun phrases with a pronominal head. Among others Hunt (1965), Scott (1988), and Ravid and Berman (2010), conclude that noun phrases become more complex in number and types of attributes, and especially post-modifiers, with age. Accordingly, to support this hypothesis, the ratio of complex lexical noun phrases, as defined in the Method section, increases significantly with age.

The fourth hypothesis takes clause constituent into consideration. A number of researchers find or hypothesize that noun phrase development differs depending on their syntactic function. Even if the development outlined above is expected to take place in all syntactic constituents to some extent, analyses of languages with a Subject–Verb–Object word order (SVO) show that subject noun phrases are, generally, shorter, simpler and more pronominal than complements (predicative phrases), objects or adverbials (Perera 1984, Scott 1988, Chafe 1994; Ravid et al. 2002). As support for the fourth hypothesis, the results in lexicality, length and complexity differ substantially depending on clause constituent (Subject, Object or Other).

Finally, the fifth hypothesis concerns itself with expert development. Kellogg (2008) maintains that writing expertise is characterized by a greater reader-awareness. Results from Johansson (2009) and Gustafson et al. (2014) show simpler syntax in the expository texts produced by adults and expert students respectively. Additionally, Gustafson et al. (2014) found a higher lexical diversity in the texts by expert students, something that in its turn may lessen the need for complex modification. Tentatively, these factors together point towards shorter and less complex noun phrases. Consequently, the fifth hypothesis is corroborated if results show shorter and less complex noun phrases in the texts produced by the expert students.

1.2 Definitions and limitations

In the present work, a noun phrase (NP) is defined as in the Grammar of the Swedish Academy, *Svenska Akademiens Grammatik* (SAG – 3 *Fraser*: Telemann et al. 1999:11-12). This is to say that the head of the noun phrase is a lexical noun, a proper noun or a nominal pronoun. The noun phrase may consist of the head exclusively, or it may have different attributes, pre- or post-positioned. The noun phrase has a nominal function, that is, its function in the clause is particularly that of subject or object or the determiner of a prepositional phrase. If the noun is put in the genitive, affixing an -s, it functions as a pre-positioned attribute to another noun phrase. For examples of lexical and pronominal noun phrases, with or without pre- or post-modification, see Table 1.

Table 1. Some examples of the basic noun phrase categories in the definition.

	single head	head with attribute/s	head with genitive attribute
lexical noun	författare 'author'	den berömda författaren 'the renowned author'	författarens namn 'the author's name'
proper noun	Dostojevskij 'Dostoyevsky'	den Dostojevskij som skrev Idioten 'the Dostoyevsky who wrote The Idiot'	Dostojevskijs årtionde 'Dostoyevsky's decade'
nominal pronoun	han 'he'		

The above definition is used to operationalize the noun phrase in a way that allows comparison between studies. Thus, it confirms closely to the ones employed in a number of previous studies, for example that of Jackendoff (1977, in Ravid & Berman, 2010): "Noun phrases as syntactic phrases consist of an obligatory head plus optional modifiers". Katharine Perera (1984:36-37) defines the noun phrase as a single lexical noun or pronoun, or a group of words with a lexical noun or pronoun as its head, a phrase that "regularly expresses the clause elements subject and object and frequently the complement" (Perera 1984:40).

It is to be understood that when investigating language development as it is manifested in the noun phrase, and within the limited scope of this thesis, the focus is not to tease apart the aspects of development which are conditioned by a general cognitive maturation on the one hand, and those which originate in practice, schooling, socialization into written text cultures

and the acquaintance with new domains of knowledge, on the other. Schooling is obviously age-related, and one may think that what we measure here is really an effect of schooling, discussed in terms of age and development. However, linguistic development in this work, is seen as the result of an iterative inter-play where cognitive maturation paves the way for the integration of knowledge and training, and vice versa. Accordingly, for practical reasons, the term *age-related development* is employed to include both of these aspects – cognitive maturation and schooling – and consequently *age* will frequently replace the more cumbersome *age and schooling*. *Expert development* or *expertise*, on the other hand, will be used for the development that ensues in adult age for those exercising writing professionally.

It is important to note that English is the language investigated by far the most and that, due to the paucity of research on Swedish, a great part of the Background section concerns itself with research carried out on English material. More to the point, this fact naturally affects the method used as well, and the definitions and categorizations used. However, the two languages are typologically close, and Swedish, like English, is a Germanic language with relatively little inflective morphology. However, even though noun phrase patterning (e.g. structures of attribution and the ordering of pre-modifiers of distinct kinds) is almost identical (Holmer, Arthur. Personal communication, June 2016), there are differences between the two languages. One of the more obvious is the definite article, which in English is an independent word – *the book* – while in Swedish it is suffixed to the noun – *boken*. In the coding for the present investigation, they would both be classified as simple lexical noun phrases. They would differ in length, though: *the book* consists of two words, *boken* of one.

In other respects, the preferred patterns of modification in the two languages appear to a great extent to be a question of frequencies. For example, the relative clause as noun phrase attribute in Swedish often correspond to non-finite constructions in English, just like *pojken som sitter i trädet* would be the most natural translation for ‘the boy sitting in the tree’, or vice versa. Furthermore, in Swedish, constructions with the genitive *-s* and compounds are frequently preferred where e.g. of-constructions are used in English. There is a strong tendency for English informative writing to use a whole new vocabulary. Words of Latinate origin are frequently used as an alternative to expanding the noun phrase for more specific reference, whereas in Swedish there is a salient tendency of compounding for the same reason. As a consequence of these distinctions and others we cannot expect the results from this study to reflect those of English exactly. (For a comprehensive analysis of a number of aspects of typological differences, see Strömquist & Verhoeven, 2004.)

2 Background

In this chapter, previous research in the topical areas is summarized. In the first two sections, 2.1 and 2.2, the studies reported are mainly of a descriptive kind, whereas in the last four sections, 2.3–2.6, noun phrase development is viewed to a great extent against the background of a cognitive paradigm within the field of language acquisition. Initially, in section 2.1 we recapitulate evidence for some well-established patterns of later syntactic development, including some critique of the tendency to equate growth in these measures with syntactic maturity. In 2.2 particular attention is paid to investigations of the noun phrase per se, and this section is followed by one in which the angle is research on subject noun phrases, 2.3. In section 2.4 we consider expert writing. Section 2.5 addresses the distinct processing conditions of the written modality in relation to the development of the noun phrase. Section 2.6, finally, concerns connections between the expository text type and the noun phrase.

2.1 Patterns in later syntactic language development

A common view on syntactic development is that most, or even all, syntactic knowledge is in place very early in the language of a young child. For example McNeill (1966) maintains that already “[a]t four, [children] are able to produce sentences of almost every conceivable syntactic type. In approximately thirty months, therefore, language is acquired, at least that part of it having to do with syntax”. Slobin, on his part (1971:74), writes that “a little child [...] masters the exceedingly complex structure of his native language in the course of a short three or four years”, and Platzack (2011) that most syntactic rules are mastered at three. Other scholars regret that the dominance of this approach has resulted in the understanding that later progress in this area is trivial and unimportant, and, as a consequence, that research on the syntactic development of grade school children, adolescents and adults is scant (Scott 1988, Myhill 2009). These scholars agree that, on the contrary, syntactic growth during the school years is substantial and vital.

However, later growth cannot be captured by the same measures as earlier development. Katherine Perera (1986b) argues that writing, which makes use of essentially other grammatical structures than speech, has to be included in a full account of language development in a literate society. In her chapter “Spoken and written syntax” (1988), Scott maintains that the grammar of older children and adolescents is indeed decisively distinct from that of younger children. Children, and later adolescents, encounter new modalities and genres – writing is added to speech, exposition to narrative – and syntax, just like vocabulary, changes significantly in new contexts. Development no longer concerns exclusively the time of the first appearance of a specific structure, but about combinations, variation and the adapted employment of old structures in new functions and contexts, and about the usage of low-frequency structures (Scott 1988).

Debra Myhill (2009) too, recognizes the relevance of studying language development after the age of eleven, and stresses the importance of analysing precisely *how* language develops into adult age, not only through the observation of the mastery of new genres, as is common,

but purely linguistically. Summarizing the research in existence, she finds that the agreement is that many linguistic constructions or units, for example the sentence, the clause and the noun phrase, increase in length and/or complexity with age. Parallel to this development there is also a decrease in the usage of personal pronouns, an increase in lexical as well as structural diversity, and in the use of the passive (Myhill, 2009).

2.1.1 Increased length of the T-unit and clause

A comprehensive overview of the relevance and reliability of diverse measures of language development in fluency, accuracy and complexity, employed in investigations of second and foreign language development in writing is given in the research report by Wolfe-Quintero, Inagaki & Kim, 1998. Most studies on later language development, however, some of which have compared spoken and written language where others have analysed only one of those two modalities, have dealt with an increase in length of the T-unit and of the clause, in number of words (Hunt 1965, O'Donnell 1967, Loban 1976, Harpin 1976, in Myhill 2009). As an overview of what development looks like in these measures, and additionally as background to noun phrase development, a brief overview of those results is given presently. A T-unit – the minimal terminable syntactic unit introduced by Hunt (1965) – is roughly equal to the sentence: a main clause with any dependent clauses attached to it. The C-unit – communicative unit – (Loban 1976), designed to provide for the specific demands of spoken language, furthermore includes non-clausal units that function independently of any clause.

In Walter Loban's longitudinal study of 211 English-speaking children (1976), a steady increase in C-unit length, for the written modality from 7.6 to 13.27 words, is found from grade three to grade twelve, and comparable numbers are found in the other studies mentioned. In Victoria Johansson's (2009) study on Swedish-speaking subjects (including the 10-, 13-, 17-year-olds and the University students of the present work, see the Participants section), mean T-unit length in written texts range from just over 9 words for written narrative texts by 10-year-olds to 16.5 words for spoken expository texts by university students. For clauses, Hunt (1965) reports that the average length increases from 6.6 words in grade four to 8.6 in grade twelve (for English), Ravid and Berman (2010) report, for written expository texts in English, ranges from 5.06 words in 9-10-year-olds to 6.84 words in 16-17-year-olds, with the corresponding numbers for Hebrew 5.81 for 9-10 year-olds to 8 for adults. For Swedish, Johansson (2009) reports a clause length of 5.7 words in written expository texts by 10-year-olds to 7.4 in written (and spoken) expositions by university students.

2.1.2 Increased use of subordination

Another feature of syntactic development, and generally viewed as a sign of writing maturity, is an increased use of subordination, calculated as the number of clauses (main clause and any dependent clauses) per T-unit. Scott (1988) reports, from the literature of English studies, a range from 1.19 for written texts by fourth-graders to 1.73 by twelfth-graders. Johansson (2009), for her Swedish subjects, reports the mean number of clauses per T-unit for all participants in her study as ranging from 1.4 for the spoken narratives produced by 10-year-olds to 2.35 for written expositions *also by 10-year-olds*. Assuming that more subordination is a characteristic of more experienced writers, this is a somewhat surprising result. Johansson explains it with the fact that 10-year-olds produce finite clauses where older language users

would employ a non-finite construction. A high rate of subordination, by this account, constitutes an early solution to the information-packaging problem, whereas longer clauses containing *inter alia* non-finite phrases, is the favoured solution among more mature writers. See the summary of Hunt (1965) below, in section 2.2 ‘The Noun Phrase in developmental research’, for a discussion of the role of the expanding noun phrase as a vital factor here.

2.1.3 Critique of the subordination and length indices

Notwithstanding the self-explanatory role of an increasing number of dependent clauses with age has been given in research, there are reservations. In Loban’s study (1976), this increase comes to a halt. Interestingly, the change is seen most pointedly in his “High (performing) Group”, occurring between the ages 13 and 17. Loban concludes that these students had developed other strategies for expressing information and the relations between thoughts in a tighter – less redundant, more packed – way. The specific structures that Loban points at are primarily non-finite constructions, but also appositives and clusters of nouns, verbs and adjectives (1976:45, 51). Furthermore, Ravid, Dromi & Kotler (2010) find that the ratio of subordination in expository texts is reduced with age and schooling (in their study, 4th-graders are compared to 7th-graders). Their interpretation is that an increasing number of other complex linguistic patterns, in particular complex noun phrases, form the basis for this change.

Scott (1988) criticizes the measure of T-unit length as an indicator of language development, in part since the growth rate is slow, but also because sentence length is at least as dependent on type of discourse as on age, and finally because longer T-units in themselves do not mean better texts. The last criticism applies also to the subordination index. Scott views the clause length index as more informative because it captures other types of complexity than subordination, for example expanding noun and verb phrases.

Like Scott, Myhill (2009) emphasizes the fact that longer and more complex structures are not synonymous to better writing in themselves. Rather, at the core of writing development is the more efficient employment of a wider set of structural and lexical tools for serving rhetorical goals, that is, the alternating use of short and simple versus long and complex structures, for variation and in different contexts. In her own comprehensive study on writing by 12-13 and 14-15-year-olds, Myhill therefor relates text quality to linguistic usage. She finds more correlation between linguistic measures and what is assessed as weak writing and good writing respectively than between age groups. Consequently, Myhill concludes as follows: “[I]t may be that age is a less helpful lens through which to view linguistic development in older writers than writing ability” (Myhill 2009:406).

2.1.4 A note on lexical development and the noun phrase

In contrast to the dominant view on syntactic growth, described in the introduction to the Background chapter, the lexical aspect of language development – an expanding vocabulary – is generally recognized as taking place throughout life. According to Nettelbladt (2007:219) a child in school age acquires around 3000 words per year, and well-educated adult English-speaking persons are said to include between 50,000 and 250,000 words in their vocabulary. Most obviously school, and later university studies or vocational education, add to the

vocabulary content words from new fields of knowledge. A hallmark of the professional writer is of course that s/he attaches great importance to the right choice of word. For instance Gustafson et al. (2014) find that the group of professional writers (identical to the Expert Students' group in the present work), showed a significantly greater lexical diversity ($p < 0.001$) than a control group of students of law and behavioural sciences. Additionally, in adolescence, writers make the acquaintance of a number of new genres, all with their own conventions and typical vocabularies, syntactic patterns and text organisations. Writing-specific knowledge like genre-knowledge, naturally, is especially salient in professional writers.

The present work does not deal with lexical development per se, but recognizes it as an important facet of noun phrase development. A specific lexical noun does not only replace the more immediately accessible, more indefinite noun or pronoun but also frequently makes unwieldy constructions with long attributes redundant. Accordingly, and highly interesting, the expanding lexicon has consequences for syntax, too.

2.2 The noun phrase in developmental research

As stated above, the expansion of the noun phrase has been subject to less attention in developmental research than the increasing length of clauses and T-units or the rate of subordination. However, at least for English, the noun phrase has been *part* of the focus (Hunt 1965, O'Donnell 1967, Loban 1976, Perera 1984, 1986b, Scott 1988, and Myhill 2009). Additionally, for English, Spanish, Dutch and Hebrew in Ravid et al. (2002), and for English and Hebrew in Ravid & Berman (2010), noun phrase development has been *the* focus of investigation, and presently a run-through of results from these studies will be made.

For Swedish, to the best of my knowledge, focus has been on the language development of younger children, and there is little, if any, research on the noun phrase from a late-development point-of-view. In Jan Einarsson's study (1978), the main perspective is socio-linguistic, even though some characteristics of the nominal in written language for upper secondary school students are contrasted with those of professional writers in this work. Lars-Olof Delsing (1993) in turn, investigated the internal structure of noun phrases in the Scandinavian languages, as did Marit Julien (2005), in the seminal work *Nominal Phrases From a Scandinavian perspective*. Ute Bohnacker analysed the nominal in early monolingual Swedish acquisition (2003). She specifically investigated the use of proper nouns with first or second person reference, the acquisition of number, gender and definiteness, and determiner omissions. Moreover, Bohnacker pointed out later development of Swedish nominals as one of the domains requiring further study (2003:248).

Einarsson's (1978) was a ground-breaking study in which written and spoken Swedish were compared from a predominantly socio-linguistic point of view, and in which nominal ratios and nominal constructions played a considerable part. The written data consisted of informative everyday prose composed by professional writers, and standardized achievement test texts written by students in year three of upper secondary school. Of interest to the present study, it was shown that the number of nouns per 100 words was significantly higher

for professional writers than for upper secondary school students (27.1 and 21.9, respectively) and for the written than for the spoken material (24.4 and 11). Reversely, the ratio of pronouns was significantly higher for the students (16.4%) than for the professionals (12.6%), and for spoken than for written texts (23.2% and 14.5%). Einarsson's conclusion was that these and other characteristics of written language were also more common in the language of professional writers than in that of high school students, in academicians than in industrial workers, and in high school students with high grades compared to students with low grades. Einarsson also compared noun phrase length in number of words, between groups and across clause constituents, and found that noun phrase objects, adverbials and complements (in that order, objects being the longest) contained more words than noun phrase subjects. We attend to subject noun phrases in developmental and cognitive research in the next section, 2.3.

We now turn to research on the noun phrase in other languages than Swedish. Investigating grammatical structures in grade four, eight and twelve, Hunt (1965) asked himself, inter alia, whether the growth in clause length could be explained by an increased use of noun modifiers, in the texts overall as well as with every single nominal head. His conclusion was that indeed it could. Children from the lowest grade used the greatest number of "simple nominals", in Hunt's account either one-word lexical nouns with or without an article, or personal pronouns. In his data, there was a high and negative correlation for clause length and the number of personal pronouns, as well as for clause length and the number of unmodified common nouns. Personal pronouns decreased dramatically from grade four to grade eight, and Hunt maintained that younger children, exactly because they wrote shorter clauses, required more personal pronouns to keep track of reference (1965:116).

Additionally, and interestingly for the present study, Hunt found that older students used decidedly more noun modifiers – adjectives, genitives, prepositional phrases, infinitives and present and past participles. He explicitly relates the increase in clause length to a phenomenon of condensation, the fact that older and abler writers tend to reduce many clauses to nouns with modifiers, consequently contributing to the higher mean number of words in the clause. He hypothesized that what older students expressed in the more concise modifiers, younger children used full finite clauses to express, as in the example "Moby Dick was a whale. The whale was very strong." [the young writer-version] versus "Moby Dick was a very strong whale." [older writer](Hunt 1965:121).

Further, to calculate noun complexity, Hunt gave one "complexity point" for each type of modifier and so counted the number of modifiers attached to each noun. The results of that analysis also led him to conclude that the noun complexity count was a valid index of linguistic maturity. He also raises the pertinent question whether depth of modification, the extent to which modifiers embed other modifiers within them, is an independent index of maturity, but does not follow that out. Finally, as Hunt takes care to point out, the fact that an increased amount of modification raises clause length in number of words, is in itself not the important improvement. What is more relevant is the greater conciseness, that is, the ability to effectively condense a higher amount of information into a lower number of clauses and words.

Loban (1976) reasons in a similar strand as Hunt, below “hypotheses being studied”: “whenever possible, a subject with high language proficiency will more frequently use phrases or non-finite constructions of all kinds in preference to subordinate clauses. This is a matter of economy: where fewer words will be as effective as many words, efficient speakers will use fewer words” (Loban 1976:18).

The conclusion of Perera’s (1984) analysis of the number of possible noun phrase pre- and post-modifications, too, is that using them, it is possible to express a variety of complex grammatical relationships in a more concise way. To illustrate the relative importance of the noun phrase in this context, she also notes that distinct kinds of writing, to a great extent, differ exactly in the kinds of noun phrases that they favour. Furthermore Perera (1984) quotes a comprehensive study by Rosenberg and Koplín (1968) where participants do not manage the entire spectrum of pre- and post-positioned modifiers until late adolescence, at 15- or 16 years of age. As salient examples of developing noun phrase types Perera mentions determiner–adjective–noun (as in *the romantic hero*), NP–preposition–NP (*the house on the moor*) and NPs in apposition (*Wuthering Heights, the house on the moor*)(1984:225).

Based on material from a variety of studies and corpora, including some of those previously listed (Hunt 1965, O’Donnell 1967, Fawcett & Perkins 1980, Perera 1984, 1986a, 1986b, Scott 1984a, 1984b), Scott (1988) concludes that noun phrases undergo a special expansion with age (in her investigation 9-19 years) and in particular contexts, specifically via post-modification: prepositional phrases, relative clauses, non-finite clauses and appositive constructions. From a developmental perspective, she writes (1988:93), pre-modification (e.g. determiner or adjectival attributes – *a house, the yellow house*) occurs before post-modification (prepositional or infinitival phrase attributes, or clausal attributes – *the house with yellow walls, the need to eat, a man who cannot read or write*).

More recently, in the article ‘Developing Noun Phrase Complexity at School Age: A Text-Embedded Cross-Linguistic Analysis’ (2010), Dorit Ravid and Ruth A. Berman report an increase with age in lexical noun phrase length in number of words. The authors view noun phrase length, calculated as scores on a scale of 1-4, as one aspect of complexity. Noun phrases produced by adults compared to all other age groups (10-, 13- and 17-year-olds) are significantly longer. This is demonstrated for English and Hebrew both. Additionally, and quite naturally, Ravid & Berman find that NPs in expository texts are longer than those in narrative texts, and that written NPs are longer than spoken ones. In this study, results also show an increase with age on a number of other scores: semantic abstractness of the head of the noun phrase, number and quality of modifiers in each noun phrase, and syntactic depth (measured as number of complex governed nodes within each NP). Apart from the age effects, there are always also interactions of modality, with NPs being consistently more abstract and more complex in written than in spoken language. Additionally, effects of genre are consistent and consequently NPs are more abstract and complex in expository than in narrative texts.

2.3 Subject noun phrases in developmental and cognitive research

In *Mind, Consciousness and Time* (1994), Wallace Chafe makes the limitations of the human mind the starting point for an examination of the interplay of characteristics and functions of the subject clause constituent. His argument is, initially, based on the production of spoken “ordinary conversation” in English, which employs SVO word order. That the attention span of individuals is limited entails that we are able to lodge no more than a restricted amount of information within working memory, frequently reproduced as the famous seven plus or minus two information chunks of Miller (1956). Applied to verbal memory, the number is held to correspond to five to seven words (Ravid & Berman, 2010:19), or the number of words that can be spoken within 2 seconds (Cowan 1992). Thus, a long noun phrase is considered psycho-linguistically complex on account of its length, considering the restrictions of verbal memory. The distance of the head of a long subject noun phrase to the verb in the clause in question, is crucial, which can be seen as a higher number of words in the subject noun phrase correlate with a higher number of mistakes in the subject–verb form congruency (Fayol, Largy & Lemaire, 1994). Roeser, Torrance and Baguely (2015) in their turn found that a complex noun phrase (in this case a coordinated NP as opposed to a simple NP) resulted in delayed lexical access. Regardless of the exact number, the limitations of verbal memory are presumed to be at the core of the preferred organisation of information in verbal communication.

According to Chafe (1994), for every sentence, the information structure-function of the subject is to express a *starting point* (or *hitching post*), by Michael Halliday in his turn, 1985, expressed as the *theme* for the thought. Taking off from, in Chafe’s and Halliday’s terminology *given* (known) information, *new* information is added only later on in the sentence, in the predicate. *Given* information, in Swedish and English, is typically expressed in a single pronoun, referring to a person or entity in the preceding sentence, whereas the *new* information introduced often demands a greater number of words and more structural complexity. As illustration we will use example (47) from Perera (1986a:101):

Last Monday the 3rd two other boys and myself did a test for the Polytechnic of Wales, building with lego bricks. *We* were given a choice, we could either build a small individual thing ourselves or build one big thing all together. [Stuart, 12)

Perera discusses the choice of the pronoun *we* (in the second sentence) as *theme* as an instance of writing competence: adhering to the convention of thematic continuity in extended discourse. But this convention combines well with a psycholinguistic explanation based on cognitive limitations. The organisation of the sentence in question enables the *writer* with a new, stable start for elaborating the heavy, new information at the end of the sentence. At the same time it may simplify for the *reader* to digest that information. Accordingly, the given-new ordering economizes on cognitive activation cost, higher for new information than for that which is accessible (present in, or closely associated to the earlier discourse), and the lowest for given information. In spoken “ordinary conversation”, according to Chafe, accessible and new information tend to be expressed with an accented full noun phrase

(1994:75). In this way, the choice of pronoun or noun depends on this given – new distinction (Chafe 1994:80).

As a consequence of cognitive economy, the occurrence of a full noun phrase subject in conversational language, especially one expressing new information is rare, according to Chafe (1994:84). He expresses this as a constraint against heavy (new and non-trivial) subjects (1994:108). *The light subject constraint*, however, is relaxed to varying degrees in different types of writing, because of the availability of more planning time in writing than in speaking. As illustration, Chafe counts 11% new subjects in the first 1000 words of an example narrative text by Hemingway, compared to the slightly more than 3% new subjects in typical conversational language. Interestingly, Chafe notes that since a pronoun is minimally informative and reliant on context cues, the employment of a full noun phrase in writing can also be seen as a reader-oriented strategy (1994:289).

We ask ourselves whether Chafe's assertion of the constraint on subject noun phrases is corroborated by the findings of other scholars. With reference to clause elements, Einarsson (1978) establishes that in his written material in total, subjects are composed by (lexical) nouns to 49.2%, and by pronouns to 44.7% (1978:143). The remaining 5.1% consist of clause subjects. The highest mean percentage of nouns, 55%, is found in the professional writers' group. Furthermore, the average length of subjects, also in the written data, is 2.3 words, the corresponding length for objects 4.2 words, for adverbials 3.3 and for complements 2.9. (These numbers include all subjects, also those that do not consist of noun phrases.) Word order in Swedish is predominantly SVO, but flexible to an extent. According to Margareta Westman (1974:155), based on a corpus consisting of newspaper articles, community information and textbooks, other than subject clause constituents are placed post-verbally in around 64% of the sentences. Consequently, it is also relevant to analyse differences depending on the varying position in the clause of the elements in question. Einarsson's results, though these are not tested for significance, tentatively show that the clause elements subject, complement and object (not adverbials) are mostly shorter *as fundamentals* (position 1) than they are overall in the spoken material, though not in the written material (1978:142). These results certainly do support Chafe's claim.

So, according to Chafe's reasoning, it entails more effort to *produce* long, or complex noun phrases in the subject constituent. Since his argument is based on English and on SVO word order, the question is whether the constituent is at issue here, or the first, leftmost position in the sentence. In Swedish an adverbial dependent clause frequently occupies this place, but other constituents may occur there as well. Incidentally, bureaucratic and formal language is characterized by "left weight" (*vänstertunga*) sentences, where frequently a great amount of information is placed in the first position. The recurring advice in writing education is to avoid this kind of sentences, because they are said to impede readability. (Bohnacker, 2007; Lundin, 2014).

In an early study by Loban (1963, in Perera 1984:100), where, still at twelve, pupils mainly use simple noun phrases as subjects, something that leads Loban to regard complexity in this

constituent as a sign of linguistic maturity. In the same vein Loban notes in his 1976 investigation that a syntactically elaborated subject (and predicate) is one of the characteristics of the High (performing) group as opposed to the Random and Low groups. Investigating noun phrase development, Scott (1988) maintains that most expansion occurs post-verbally, i.e., in other than subject position. We observe that all of these conclusions are based on English.

When Perera (1986b) investigates to what extent nine- and twelve-year-olds are able to differentiate speech from writing, she highlights the complex noun phrase subject as one of the writing-typical grammatical constructions. In Perera's study a complex NP is defined as anything other than a pronoun, a proper noun or a determiner plus a noun. She also categorizes noun phrases with "notable complexity", that is, anything more than [D (determiner) Adj N], [D N prep N] or two co-ordinated simple NPs. Perera finds a steady growth in complexity, and concludes that grammatical function is indeed critical, since children are well able to use longer and more complex noun phrases as objects, i.e. post-verbally. Perera believes that the reason why the complex (and long) NP subject is a late acquired structure is the increased distance from head noun to verb, and the additional possibility of other nouns in between, vying for the interpretation as subject (1984:292). These structures place a heavy burden on short-term memory, she writes.

Ravid, van Hell, Rosado and Zamora (2002) investigate all subject noun phrases across the four variables *age* (9 to 11-year-olds versus university graduate adults), *genre* (narratives versus expository texts), *modality* (speech versus writing) and *language* (Dutch, English, Hebrew and Spanish). Noun phrases are categorized as pronominal or lexical and pronominal subjects are divided into personal or impersonal. Furthermore, the lexical subjects are classified as simple or complex (governing one single versus more than one lexical noun). The results from the variable age, along with some typological observations, are the most relevant ones for the present work, but we return to modality and text type in 2.5. Results show that the percentage of lexical subject NPs increase as a function of age in Dutch, English and Hebrew, but not in Spanish. In adult written expository discourse the percentage of lexical subjects amount to about 30 % (Ravid et al. 2002). The percentage of complex lexical NPs increases with age in Hebrew and Spanish, but not in English. Incidentally, this could be an effect of the parallel Latinate vocabulary in English, mentioned in the Introduction as a typological characteristic, and decreasing the demand for more complex noun phrases.

Ravid et al. (2002) interpret the more frequent occurrence of lexical and more complex noun phrases as subjects in the adults' texts as a result of adults' better ability "to represent and process information and to encode it in a densely packaged form" (2002:20). Referring to Chafe (1994), Du Bois (1987) and Clark & Wasow (1998), the subject position in SVO languages is pointed out as "particularly difficult to realize [...] with highly lexical and complex NPs" (2002:20).

2.4 The development of writing expertise

Only few researchers have investigated the area of writing expertise, and even fewer have concerned themselves with the syntactic structures employed by expert writers. Hunt (1970) concludes that “the average twelfth-grade writer has achieved full syntactic maturity *unless that student continues on a path requiring educational and occupational practice in text writing*” (emphasis added). Scott (1988), problematizes the concept of one adult model of writing competence and argues, firstly, that we need to know more about adult competence, and secondly, that we need to allow for a broad span of competencies. Kroll (1981, in Perera 1986a) sketches grammatical development in writing in four phases, and suggests that the last phase, “integration”, when the writer is able to express her/his own style and voice, is reached only by a limited number of writers. Myhill (2009), in her turn, views the competent writer as a designer, confidently familiar with the repertoire of linguistic alternatives, and furthermore with how to make those choices that are rhetorically efficient. Finally, graphical rhythm, lexicon and syntax are parts of Lisa Holm’s investigation of ten contemporary Swedish novels (Holm, L. Svensk romanprosa, in progress).

Kellogg (2008) is, as far as we know, the one who has concerned himself most with the development of writing expertise, though not with syntactic patterns per se. He builds on the two seminal writing stages-cum-strategies formulated by Bereiter & Scardamalia (1987) as *knowledge telling* and *knowledge transforming* and complete them with a third stage, *knowledge-crafting*. The model builds on the fact that limitations of the cognitive resources available to the writer play a big part in writing development, as developed by Deborah McCutchen (1996, 2000, 2011).

According to Bereiter & Scardamalia’s model, a writer using the knowledge-telling strategy is overwhelmingly concerned with putting on paper everything that she knows about the topic in question, in the order that it comes to mind. Pertinently, it has been called a “memory dump” (Pea & Kurland 1987:293, in Myhill, 2009). The result may be characterized by a lack of structure and the absence of rhetorical design. Even if the writer is to some extent, or at some phase, aware of the reader, she is not able to realize this awareness because other components of the complex writing process, like for example spelling, demand her attention.

A knowledge-transformer, on the contrary, is able to work the text, planning and revising it, considering vocabulary and linguistic choices in order to make the text express what she wants it to say. Kellogg’s addition, knowledge-crafting, is the stage where the writer successfully handles both her intentions, the text itself and a representation of the text as it is perceived by the reader, simultaneously. The knowledge-crafter has practised text composition to such an extent that some of the many components work automatically, which makes it possible to do several things at the same time. The most salient characteristic of the knowledge-crafter is reader-orientation. Kellogg emphasizes the fact that whereas a writer may command the knowledge-transforming strategy after a period of training of ten years, at least ten more years are needed to become an expert.

2.5 The written modality and the noun phrase

The focus of the present work is not to contrast the noun phrases in speech with those in writing, nor is it a comparison between the expository and other genres. However, before we proceed to the investigation proper, we are going to see how, in writing, some favourable conditions for longer, more lexical and complex noun phrases are in place, at the same time as the demand for these structures arises in the informative expository text.

A number of scholars explore and describe the differences between the two modalities writing and speaking. These differences concern organisation, lexicon and syntax (Halliday, 1985; Biber, 1988; Chafe, 1994; Strömquist, 2009:94). Most accounts depict the process of writing as comparatively lasting, slow, reflected and independent of context, and the output language of writing as more precise, concise, elaborate and explicit than that of speaking. However, especially Biber (1988) problematizes a simple dichotomy view on speaking and writing, emphasizing that the differences can be seen more to the point as a continuum and in this way making allowances for numerous and distinct text types. Halliday, in his turn, writes of “clusterings of particular tendencies” (Halliday 1985:45). One of the crucial tendencies is a higher lexical density in writing – the proportion of lexical or content words as compared to function (grammatical) words. Relevant to the aim of the present study of the noun phrase in expository texts, a higher lexical density results in a higher amount of information. In Halliday’s description the powerful linguistic inclination is to express information in nominals. Halliday also puts forward a seductive argument about how speaking and writing leads to different views of reality where writing expresses a world populated by “things” and speaking one of “happening”: dominated by nominals and verbs respectively.

For our aim, one of the most relevant characteristics of written discourse is that it stands to a higher extent un-supported by the context – the here and the now, and the immediately present interlocutor. The effect is that for unhindered communication, for instance deictic pronouns and adverbials of place must be replaced by more precise and elaborate expressions. Perera (1986b), writes that in writing, the function of more complex structures is also to obtain less redundancy and more variation. This is a variation badly needed since the prosodic and expressive features of speech are not available in writing, features that are frequently used, for example, to mark information structure.

As we account for above, one of the more elaborated structures typical of writing that Perera (1986b) investigates is the complex noun phrase in the subject function (summarized in section 2.3). Perera notes, for instance, that in speech the reiteration of a pronominal theme is unproblematic, but for a writer variation, thematic continuity and an efficacious positioning of focus is vital. Perera goes so far as to say that different types of writing are to a great extent recognizable by the types of noun phrases utilized.

Perera (1986b) investigates children’s ability to distinguish spoken language from that of writing. When comparing her own data with data for adults from Quirk et al. (1985) she notes that already at nine, children use a higher percentage of complex subjects in writing (8-9%), than do adult subjects in informal speech (7%). For twelve-year-olds the corresponding

percentage (of complex subjects in writing) is 12%, and for adult “serious writing” 16%. In adult scientific writing the ratio is 38%.

2.6 The expository text type and the noun phrase

Written expository discourse, the *macrogenre* (Grabe 2002, in Berman and Nir 2010) of informational text, is at the farther end of Biber’s continuum of spoken to written language. The reason is to be found in the basic function of this text type: to communicate information. In doing so it is distinct – in organization, content, vocabulary and linguistic structures – from other text types, write Berman & Nir (2010), who furthermore define it (2010:101) as “[containing] at least one idea, opinion or claim”. The prototypical language of an expository text is impersonal, informative, complex, abstract and dense. Furthermore Ravid, Dromi & Kotler (2010:141) assert that “expository discourse fosters grammatical complexity from early on”. As we can see, the expository text is a potentially rewarding place to look at noun phrases; we might reasonably expect to find these structures utilized to their information packaging limits here.

It is a well-known fact that the expository genre presents a radical challenge for school children when they first encounter it. So much so, in fact, that the phrase “the fourth grade slump” denominates the drastic decline in reading comprehension that has been observed in school children in the U.S.A. Snyder & Caccamise (2010) account for the constraints and difficulties that hinder comprehension, relevant to the present work even though we deal with the production rather than the comprehension of expositories.

Berman and Nir (2010) conduct a developmental investigation on the English and Hebrew data collected within the Spencer project (as for four of the participant groups in the current study). They find that, due to the preference in expository texts for general statements and an impersonal stance, expletive subjects and impersonal and generic pronouns are frequently used. They specifically report a heavy reliance for the youngest writers on the generic *you*, whereas older writers prefer to use an abstract subject, often of a certain length and complexity (2010:114). Impersonal pronouns also stand in as subjects for entire propositions (“it is [...] to imagine a world of conflict”). Along with this tendency, writers employ a great number of long and complex noun phrases as “frames of content”, typically using non-finites to serve as complements to nouns. Not only do the written expositions in Berman & Nir’s study display a higher lexical density, they also favour long, infrequent and (for English) morphologically complex Latinate origin-words.

Furthermore, comparing genre and modality, Ravid et al. (2002) find that the increase in lexicality is more substantial in expository texts than in narratives, in written texts than in spoken ones, and in the written expositions when compared to the written narratives. Moreover, the written expositions contain a higher percentage of complex lexical NPs than the written narratives, and also than the spoken expositories. The same characteristics (a higher NP lexicality and complexity in written expository texts as opposed to narratives) are explained by the fact that this genre is concerned more with “abstract concepts, processes and

ideas”, and they occur in the written modality more than the spoken one because on-line processing constraints in the spoken modality are cancelled when writing.

It appears that both the relative context-independence of the written modality and the informative intent (and content) of expository texts create a clear demand for longer, more lexical, and complex noun phrases. At the same time, the conditions of production of writing allow for these constructions, providing the time needed for deliberation in word choice as well as in longer and more complex constructions.

3 Method

The data for the present study was derived from three existing corpora, which are introduced in this chapter. Further, the participants and procedure in the original experimental studies are described, together with the ethical considerations made at the time of data collection. We also take an introductory look at the material in the current investigation. After that the coding, and the rationale for the coding, are explained. Finally some technical information about the analyses is given.

3.1 Corpora

This study is based on the corpora from three previous projects: “The Spencer Project” (2002), “Dynamics of Perception and Production in Text Writing” (2008), and “Expert Writing” (in progress). The experimental data for the 10-, 13-, 17-year-olds and adult university students in this study were collected for “Developing literacy in different contexts and different languages”, the “Spencer project” (Berman & Verhoeven, 2002). For a thoroughgoing description of the Swedish data subset from this project, see Johansson (2009). With the goal of studying age-related development of text production in the written and spoken modalities during the school years, in seven different languages, the Spencer project was cross-linguistic, with Ruth Berman as its main coordinator.

Data for the 15-year-olds originate from the project "Dynamics of Perception and Production in Text Writing" (Johansson et al. 2008). The main object was to explore the interplay between perception and production during text writing. The participants consisted of four groups: 15-year-olds with and without reading and writing difficulties, and adults with and without the same difficulties. However, in the present study, only 15-year-olds from the control group participated.

The adult expert students derive from “Expert Writing - divine inspiration or hard work?” (Johansson, V. Expert writing – divine inspiration or hard work? Swedish Research Council: VR2011–2290, in progress). Participants in this project were recruited from a creative writing program at university, to which admission was gained through an entrance test demonstrating writing skills. Age range in this group was wide, 21–68 years. In Expert Writing the main focus is on analysing how good writers develop writing expertise during the course of their training, rather than on enrolling a group of participants with a homogeneous background with respect to age and previous education.

3.2 Participants, ethical issues and procedure in the original studies

Participants in the original experimental studies were distributed into six groups: 10-, 13-, 15-, 17-year-olds and two adult groups: university students and expert students. To ensure comparability between subjects from the different age groups, to the extent possible, and to minimize influence from extraneous variables, participants in all studies were monolingual native speakers of Swedish, without reading or writing difficulties, and from well-educated backgrounds and well-established schools. Because the methods and focus of teaching along

with computer literacy were considered by the investigators to have the greatest influence on the results, the younger participants were, as far as possible, chosen from the same class or at least from the same school. Of the university students 50% came from the humanities and social studies and 50% from the natural science studies.

For the younger Spencer participants, the 10- and 13-year-olds, teachers filled in a questionnaire rating reading and writing abilities. 15-year-olds in *Dynamics of Perception and Production in Text Writing* were screened for reading and writing difficulties, and for the Expert Students, data were supplemented by in-depth interviews about reading and writing habits, and a working memory test was carried out. All participants gave their informed consent in writing, and moreover, for 10- and 13-year-olds, parental consent was given. See Johansson (2009) for more information.

In all three of the above studies, the procedure followed the same main outline. Participants knew that they were taking part in a research project on “how people in different ages write” and were asked to do their very best. They watched the same short elicitation movie about unresolved conflicts in a school setting – cheating, mobbing, theft and vandalism, and wrote one narrative and one expository text. In the Spencer project, participants additionally performed the corresponding oral tasks: one narrative and one expository oral text, all on the topic of problems between people. Even if the other task components differed somewhat between the projects, there is no reason to believe that the texts used in this study were substantially affected by those differences. The instructions for the written expository text were identical across projects. For a more extensive account of the experimental procedure, see Johansson (2009).

3.3 Overview of participants and data in the current study

3.3.1 Participants

Comparing age groups based on cross-sectional data corresponds to a between-subjects design. To balance the groups for the present study, the 16 texts by the whole group of four male and twelve female expert students were included. Furthermore, the texts by 16 participants, eight female and eight male, were randomly chosen, though evenly distributed over task order, out of the original 20 participants in each of the age groups 10-, 13-, 15-, 17-year-olds and graduate school university students. This measure was taken since the task order variable was the one yielding most significant differences in Johansson (2009). Groups of 16, moreover, are large enough for parametrical statistics, and limited enough to make it possible to manually code all noun phrases. An overview of numbers, gender, age ranges and original projects for the participants in the present study is shown in Table 2.

Table 2. Overview of participants: for each group number, gender, age range (years:months) and project.

	number	gender	age range	project
10 years	N=16	Female N=8 Male N=8	10:2-11:2	Spencer Project
13 years	N=16	Female N=8 Male N=8	13:2-14:1	Spencer Project
15 years	N=16	Female N=8 Male N=8	13:5-15:10	Dynamics in Text Writing
17 years	N=16	Female N=8 Male N=8	16:11-18:3	Spencer Project
University Students	N=16	Female N=8 Male N=8	23:11-43:7	Spencer Project
Expert Students	N=16	Female N=12 Male N=4	21-68	Expert Writing

3.3.2 Material

In all three original studies, as described above (section 3.2), the procedure was the same. Relevant to the present thesis participants were asked to write an essay on the topic of problems between people, discussing these problems and possible solutions to them. The researchers were receivers of the texts. Writing time was about 30 minutes. The written texts were collected by means of keystroke logging providing the opportunity to study the on-line writing processes (Strömqvist et al. 2006). However, the object of this study is the final product: the finished text.

As can be seen in Table 3, the material consisted of 96 expository texts written by equally many participants distributed on six different groups of age and expertise. It is appropriate to make a special note that the material is uniquely comparable. A great number of the developmental studies referred to in the Background section compare texts in different genres and from varying writing tasks, collected in different conditions, for different age groups (Hunt, 1965; Loban 1976; Einarsson 1978; Perera, 1984; 1986b). In the present work, as stated, all participants in all groups wrote the same type of text in the corresponding experimental conditions.

Table 3. Overview of material: for each group number of participants, total, mean, median text length in number of words, minimum and maximum values and standard deviation.

group	no	text length, total	text length, mean	text length, median	text length, min-max	sd
10 years	N=16	1902	119	123	41-178	<i>42.3</i>
13 years	N=16	3748	234	221	72-484	<i>107.1</i>
15 years	N=16	7438	465	495	209-764	<i>158.9</i>
17 years	N=16	8694	543	542	256-836	<i>140.0</i>
University Students	N=16	7422	464	497	187-752	<i>164.2</i>
Expert Students	N=16	7613	476	437	251-817	<i>182.7</i>
Total (whole material)	N=96	36817	384	384	41-836	<i>205.4</i>

An introductory look at the material shows that mean text length nearly doubles from 10- to 13-year-olds ($M=119$, $M=234$). No doubt, a greater number of noun phrases can be expected in the longer texts, and consequently the possibility of greater variation in the structure of noun phrases in these texts. The texts produced by the groups originating in the Spencer project (10-, 13-, 17-year-olds and University Students) were analysed by Johansson (2009), who reported age effects on text length, and who also observed a developmental leap between 13- and 17-year-olds. In the present study the group of 15-year-olds are added, and a substantial increase in text length is shown between 13- and 15-year-olds ($M=234$, $M=465$). There is also an increase between 15- and 17-year-olds ($M=465$, $M=543$), but subsequently text length in fact *decreases* between 17-year-olds and the two adult groups ($M=464$, $M=476$). The decrease from 17-year-olds to the University Students was reported already in Johansson (2009), and it is highly interesting to see that this effect is reinforced in the Expert Students. We can see that the median value is quite close to the mean in most cases. An ANOVA shows a significant effect of group on text length ($F(5, 90) = 22.606$, $p = 0.000$). An increase in text length with age is in concordance with earlier research, for example O'Donnell (1967), who reports this fact as one of his principal findings. The results for text length – medians, range and outliers – are displayed in Figure 3.

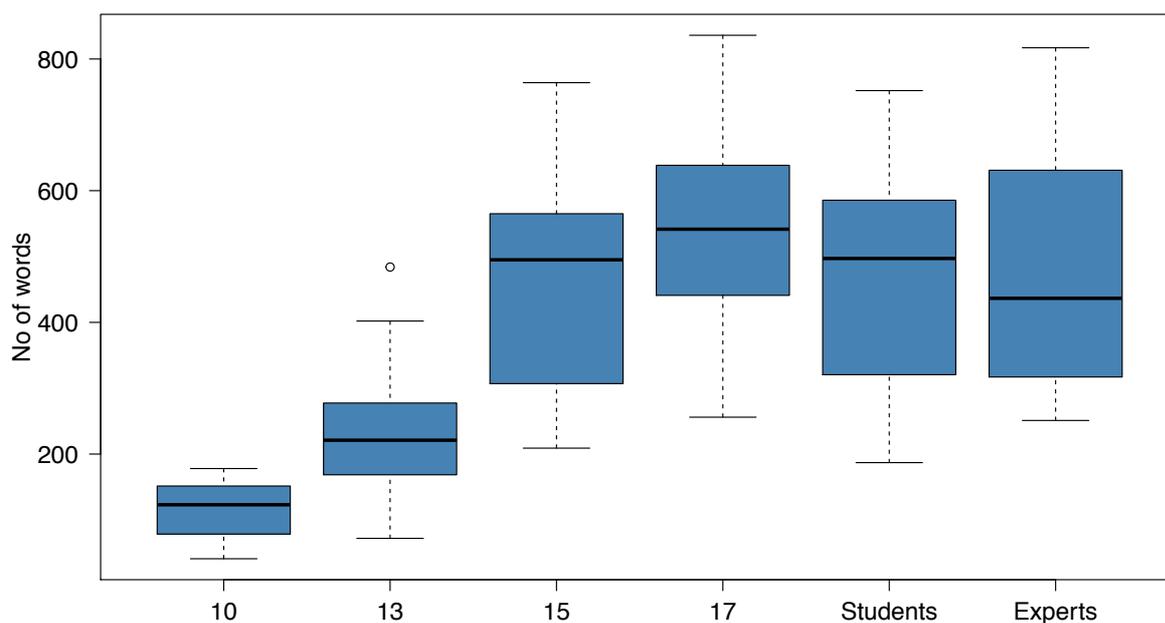


Figure 3. Median, full range and inter-quartile range for text length in number of words, across groups.

As stated in section 1.2 above, a noun phrase in this work is defined as a single lexical noun or pronoun, or a group of words with a noun or pronoun as its head (SAG 1999, Jackendoff 1977, Perera 1984:36-37). In the material in question, consisting of 36,817 words, 8,670 noun phrases were found (Table 4).

Table 4. Overview of noun phrases: for each group total number, mean, median, minimum – maximum values, standard deviation and correlation text length – number of noun phrases.

group	no of NPs, total	no of NPs, mean	no of NPs, median	no of NPs, min-max	sd	correlation words-NPs
10 years	467	29	32	14-43	9.2	0.94
13 years	933	58	56	22-121	25.9	0.98
15 years	1799	112	121	52-181	37.3	0.98
17years	2052	128	125	55-203	34.3	0.97
Students	1614	101	98	44-179	36.9	0.14
Experts	1805	113	112	54-202	49.3	0.94
total material	8670	90	84	14-203	48.5	0.90

Interestingly, for all groups except for the University Students, who show a saliently low correlation on this count, the mean number of noun phrases correlates highly with the number of words in the texts by each age group. The University Students show both a lower mean number of NPs and lower minimum and maximum values than do both 17-year-olds and Expert Students. Assuming that the writers in the University Students group convey in their texts (at least) as much information as do 17-year-olds and Expert Students, the result may indicate that these writers tend to pack information into fewer but more expanded NPs rather than into a higher number of less expanded NPs, or into more specific ones.

However, from these figures we are not able to deduce the actual differences in the percentage of pronouns compared to that of lexical nouns, or developmental changes in length and complexity. This is exactly what we intend to find out more about.

3.4 Coding

All noun phrases in the 96 texts (N=8670) were identified and classified. In the following section, the coding categories for noun phrase lexicality, length and clause constituent as well as for complexity will be listed and described.

3.4.1 Lexical and pronominal noun phrases

The basic distinction made in this work is the one between lexical noun phrases (*house, parents, Alison*) and pronominal ones (*it, them, she*). See Table 5 for examples. Following Einarsson (1978:77), the pronoun group in this work consists of those pronouns that function as the head of a noun phrase (see the definition of the Swedish Academy Grammar, in section 1.2 Definitions and limitations section). Pronouns that modify lexical nouns are classified as adjectival modifiers. Reflexive pronouns are not classified as noun phrases but are seen as parts of the verb phrase (following Lundin 2014). Pronouns are a highly heterogeneous group of function words, covering a number of distinct references, and subsequently different researchers have divided them into varying categories. For instance Ravid et al. (2002) utilized the distinction *personal pronouns*, meaning those of the first and second person, to draw conclusions about discourse stance. Hunt (1965) separated out personal pronouns, too, but included pronouns of all persons, excluding only expletive *it*. In the present study, merely the distinction between lexical and pronominal noun phrases is maintained.

Table 5. Coding of noun phrases: simple or complex, length, constituent and complex or simple.

*När **det**¹ rör sig om **skolmiljön**² blir **lärarna**³ **det naturliga centrat för problemlösning**⁴ (wu19mDEW)*
 ‘When **it**¹ concerns **the school environment**² **teachers**³ become **the natural centre of problem solving**⁴’

noun phrase	lexicity	NP length	constituent	complex/ simple NP
<i>det</i> (‘it’)	pronoun	1	subject	–
<i>skolmiljön</i> (‘the school environment’)	lexical	1	other	complex
<i>lärarna</i> (‘the teachers’)	lexical	1	subject	simple
<i>det naturliga centrat för problem-lösning</i> (‘the natural centre for problem solving’)	lexical	5	other	complex
<i>den grupp som mobbaren förhåller sig till genom sitt mobbande</i> (‘the group that the bully relates to through his/her bullying’)	lexical	10		complex

Noun phrases in lexicalised adverbs wherein the noun part has lost its original nominal reference, like *idag* (‘today’), *istället* (‘instead’), *ifall* (‘if’), often written or possible to write as one word, were not categorized as noun phrases. Likewise, it is to the point to state that in the present analysis, the lexical noun group include proper names.

3.4.2 Noun phrase length

A longer noun phrase (in number of words) is not necessarily the equivalent of a more complex one. Still, we get a complementary estimation of complexity when calculating noun phrase length, following Ravid & Berman (2010). A long NP typically includes either a clausal attribute, and/or a combination of attributes, or coordinated attributes. With regard to the noun phrase length measure it is appropriate to say that the number of words counted are the graphic words. We find a few samples in Table 5.

3.4.3 Simple and complex noun phrases

The next distinction in this work is the one between simple and complex noun phrases (*huset*: ‘the house’ – *tvåvåningsbyggnaden med balkonger på den soliga sidan*: ‘the two-storey building with balconies on the sunny side’). In the present work, structures appearing later in development are considered more complex. Scholars like Scott (1988), investigating English, have emphasized post-positioned attributes and frequently the prepositional phrase as appearing and/or becoming more frequent later in development. Perera (1986b) defines a pronoun, a proper noun or a determiner plus a noun as simple and “anything else” as complex. However, she also categorizes noun phrases with “notable complexity”, that is, anything more than [D (determiner) Adj N], [D N prep N] or two co-ordinated simple noun phrases.

Since the current study is concerned with Swedish, the definition of complexity must be tentative. The category here includes compounds consisting of two lexical nouns, and lexical nouns modified by a noun in the genitive, a prepositional phrase, an infinitival phrase, a

restrictive relative clause or other clausal attributes. (The restrictive relative clause is “a necessary modification, delimiting the set of referents referred to by the relative head” (Platzack, 2003:268).) This is to say that complex NPs are identical to NPs with post-positioned modifiers, with the addition of genitives and compounds. Appositions, predicative phrases (complements) and non-restrictive relative clauses are not counted as part of the noun phrase, following Einarsson (1978).

Furthermore, a noun phrase is delimited at its uttermost boundary (following Ravid & Berman 2010), so as to avoid double counts. It is then categorized in accordance to its most complex modifier, in the order compounds > genitives > prepositional phrases > infinitival phrases > restrictive relative clauses > other clausal modifications. For an example, we turn once again to Table 5. Consequently, the last sample: *den grupp som mobbaren förhåller sig till genom sitt mobbande* (‘the group that the bully relates to through his/her bullying’), is categorized as relative clausal even though it also contains a determiner and a prepositional phrase. On the other hand, single lexical nouns without a modifier, with a determinative or adjectival modifier (numerals, possessive pronouns, adjectival pronouns and adjectives), or other pre-positioned attributes, together with compounds consisting of no more than one noun were counted as simple lexical noun phrases (Lundin 2014).

3.4.4 Clause constituent

As we have seen in the Background section, several scholars point out distinctions in the developmental patterns of noun phrase lexicality, length and complexity depending on the syntactic functions and/or position of the noun phrase (Einarsson, 1978; Perera 1986b; Scott 1988; Chafe 1994; Ravid et al., 2002). Consequently, noun phrase categories in the current work further comprise Subject, Object or Other constituent. The Subject category includes subjects in passive clauses. Correspondingly, the Object category consists of direct and indirect objects, and Other of complements, prepositional objects (see Hultman, 1987) and adverbials.

3.5 Coding and analyses

The 96 texts were converted and transcribed into the CHAT format (MacWhinney, 2015), and all noun phrases (N=8670) were identified. Single lexical nouns, pronouns, and groups of words with a noun or pronoun as their head – all were coded on three different levels. They were categorized as 1) lexical or pronominal, 2) simple or complex, and finally 3) according to clause constituent as Subject, Object or Other. Subsequently data on number of noun phrase tokens, frequencies for various combinations of the distinct categories, and lists of the actual phrases were obtained in the CLAN program *FREQ* (MacWhinney, 2015), a powerful and well-established tool for the analysis of linguistic development. Additionally, noun phrase length in number of words was calculated in R (2014), using an R script helpfully developed by Joost van de Weijer.

Moreover, statistical computations were made in R (2014). An ANOVA was carried out initially, to find group effects on lexicality, length and complexity respectively, overall as well as separately for Subject, Object and Other clause constituents. Post-hoc comparisons

were made with 10-year-olds as intercepts, using contrasts analyses and Tukey's HSD procedure to find out more on significant variation between groups. Correlations were calculated in R and tested with Pearson's correlation coefficient.

4 Results

Results are organized in sections for the three aspects of lexicality, length and complexity respectively. In each of the sections, we start by looking at the development of noun phrases in all clause constituents together, and then the three clause constituent categories are investigated separately. Subsequently, for each aspect, a discussion is held with the help of some relevant examples from the different groups. Following the Results section there is a general, concluding discussion going out from the two distinct developments attended to in the current work.

4.1 Lexicality – pronouns and lexical nouns

To test the first of the hypotheses, lexicality was calculated as the percentage of lexical noun phrases out of all noun phrases (i.e. lexical and pronominal), for each of the six groups. This measure was completed with the ratio of pronominal NPs out of all NPs, following the procedure of Ravid et al. (2002). The same procedure was used firstly for noun phrases overall, and hereafter, to test the fourth hypothesis, lexicality was analysed for Subject, Object and Other constituents separately.

4.1.1 Lexicality overall – quantitative analysis

The total number of noun phrases, those headed by a lexical noun as well as by a pronoun, and the number in the two categories separately was obtained in CLAN (the FREQ program). The percentage of each of the two categories out of all noun phrases, in all syntactic functions, was calculated for each of the six groups (lexical NPs/all NPs, pronominal NPs/all NPs) and the results can be viewed in Table 6.

Table 6. Mean percentage noun phrases with a lexical head and with a pronominal head, across age groups. Minimum and maximum values within brackets.

group	lexical NPs	pronominal NPs
10 years	42.6% (7-53%)	57.4% (47-93%)
13 years	38.3% (26-62%)	61.7% (38-74%)
15 years	44.4% (29-56%)	55.6% (44-71%)
17 years	53.8% (42-67%)	46.2% (33-58%)
University Students	62.2% (41-80%)	37.8% (20-59%)
Expert Students	60.5% (35-78%)	39.5% (22-65%)
total	50.3% (7-80%)	49.7% (20-93%)

The ANOVA showed a significant effect of group on the proportion of noun phrases with a lexical head out of all noun phrases, ($F(5, 90) = 14.15, p = 0.000$). Post-hoc analyses yielded significant differences between 10-year-olds and 17-year-olds, $p = 0.004$, as well as between 10-year-olds and University Students on the one hand, and 10-year-olds and Expert Students on the other, both $p = 0.000$. Moreover, significant effects were found between 13-year-olds

and the following groups: 17-year-olds ($p = 0.001$), University Students ($p = 0.000$) and Expert Students ($p = 0.000$). Also for 15-year-olds differences to all three of the oldest age groups 17-year-olds ($p = 0.014$), to University Students ($p = 0.000$), and to Expert Students ($p = 0.001$) were found to be significant. Furthermore, significant differences between 17-year-olds and University Students, ($p = 0.027$), were found. The results are additionally displayed in Figure 4, where the plot to the left shows lexical NPs. In the right-hand part of Figure 4, the (reverse) ratios for pronominal NPs can be seen. Here it is also made apparent that individual variation was greater for example within the two adult groups than in the 10-year-olds, 15-year-olds and the 17-year-olds. Note also, firstly, that even though there was a clear increase in the proportion of lexical noun phrases, and reversely, a clear decrease in the proportion of pronominal NPs from the three youngest age groups to the three older groups, the lowest proportion of lexical NPs was not found for the 10-year-olds, the youngest group, but for the 13-year-olds. Secondly, that the proportion was lower in the Expert Student group than in the University Student group, though not significantly so.

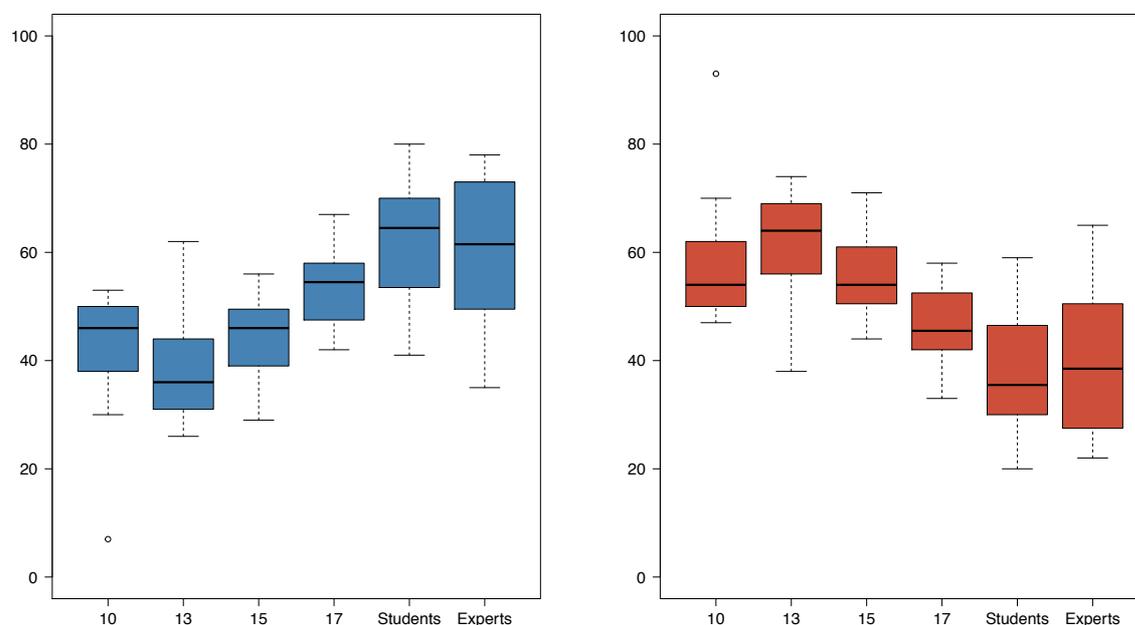


Figure 4. Mean percentage lexical NPs to the left, pronominal NPs to the right, out of all NPs.

4.1.2 Lexicality across clause constituent – quantitative analysis

To find out whether there were group effects on the proportions of lexical and pronominal noun phrases for the distinct clause constituents, the same calculations as for all noun phrases were subsequently made for NPs in the Subject, Object and Other function categories separately. Numbers of all NPs as well as for lexical and pronominal NPs in Subject, Object and Other functions respectively were obtained in CLAN (FREQ) and the ratio of each of the two categories out of all NPs, in Subject, Object and Other function, was calculated for each of the groups (pronominal NPs/all NPs, lexical NPs/all NPs). The mean percentage of lexical as well as pronominal NPs, for each group and for each syntactic function, can be viewed in Table 7.

Table 7. Mean percentage lexical noun phrases out of all (lexical and pronominal), across clause constituent and group.

	10 years	13 years	15 years	17 years	University Students	Expert Students	total
lexical NP as Subject	19.3%	16.6%	23.1%	28.4%	41.6%	35.8%	27.5%
lexical NP as Object	64.9%	60.1%	61.1%	75.7%	70.9%	75.8%	68.1%
lexical NPs as Other	73.5%	71.1%	72.4%	85.7%	87.8%	86.9%	79.6%

The same results are further illustrated in Figure 5. Scrutinizing the distribution of lexical noun phrases across different clause constituents several results stand out: Firstly, that lexical noun phrases in all age groups were decisively more common as Objects (group means ranging from 60.1% - 75.8%) and as Other constituents (71.1% - 87.8%) than as Subjects (16.6% - 41.6%). And secondly, that the increase from the lowest number to the highest in the proportion of lexical NPs was clearly more substantial in Subject function than in Object and Other functions. The highest proportion, found among the University Students (41.6%), was 2.5 times higher than the lowest one, found in the 13-year-old group (16.6%). A third interesting observation was that, mirroring the counts for lexical noun phrases overall, it was true also for each and all three clause constituents that the lowest proportion of lexical NPs was found in the *next* youngest age group, the 13-year-olds, and not in the youngest, the 10-year-olds. Finally, the mean percentage of lexical NPs in Subject constituent was lower in the Expert Students group than in the University Student group, whereas for Object noun phrases it was higher for the Expert Students.

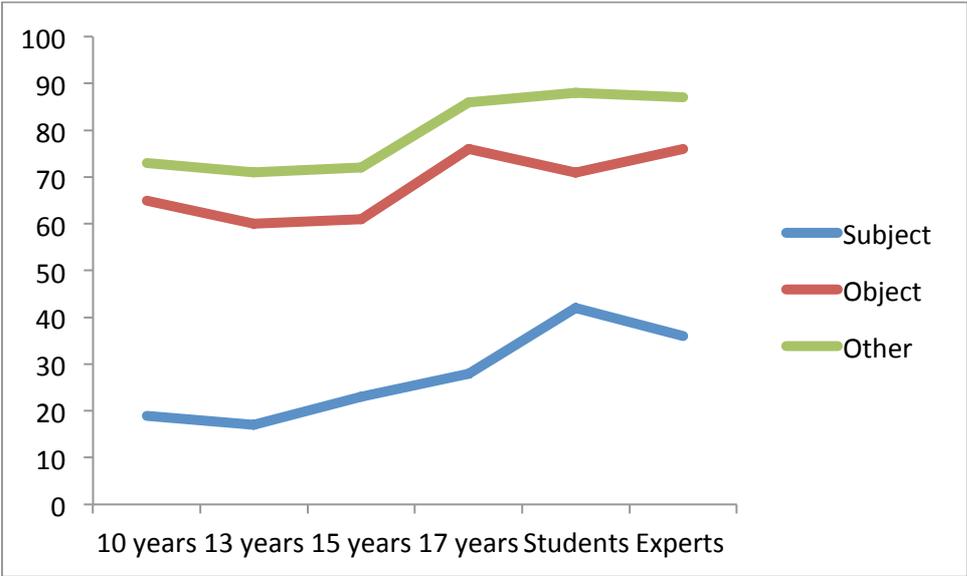


Figure 5. Percentage lexical noun phrases across clause constituent.

4.1.3 Subject noun phrase lexicity – quantitative analysis

Testing the significance of differences in the proportions listed in Table 7, and starting with lexicity in noun phrases as Subjects, a significant group effect of on the proportion of lexical noun phrases in the Subject constituent ($F(5, 90) = 10.00, p = 0.000$) was found. Post-hoc analyses revealed significant differences between 10-year-olds and 17-year-olds ($p = 0.039$), between 10-year-olds and University Students ($p = 0.000$), and between 10-year-olds and Expert Students ($p = 0.000$). Significant differences were also found between 13-year-olds and University Students ($p = 0.000$), between 13-year-olds and Expert Students ($p = 0.000$) as well as between 15-year-olds and the two adult groups, University Students ($p = 0.001$) and Expert Students ($p = 0.052$). Furthermore, a significant difference between 17-year-olds and University Students ($p = 0.003$) was found. An overview of the proportion of lexical nouns and pronouns in Subject function in the different groups can be seen in Figure 6.

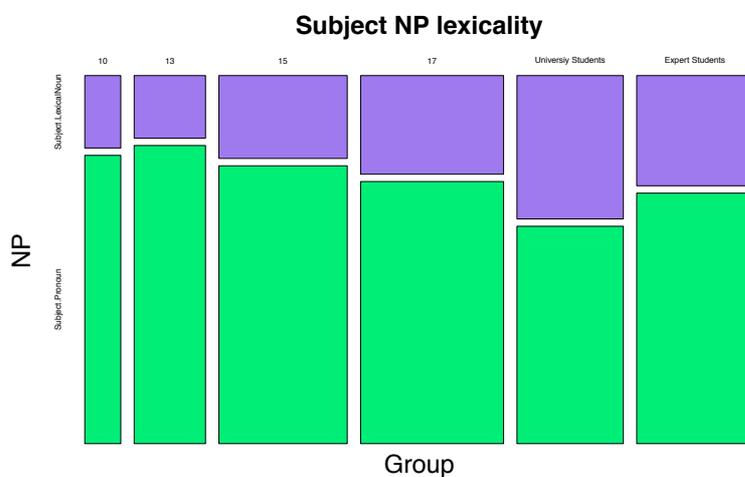


Figure 6. Proportion of lexical (purple) and pronominal (green) noun phrases in the Subject constituent.

This type of graph also visualises the fact that the raw number of lexical nouns in the Subject function differed decidedly between groups, and that 10-year-olds, with the lowest raw number, produced only slightly more than one sixth of those of the University Students, the group with the highest number.

4.1.4 Object noun phrase lexicity – quantitative analysis

The analysis of group effects in noun phrase lexicity in Object function yielded a significant effect of group on the percentage of lexical noun phrases in this clause constituent ($F(5, 90) = 3.25, p = 0.010$). Post-hoc tests further showed a significant difference between 15-year-olds and 17-year-olds ($p = 0.010$). As can be seen in Figure 7, proportions were comparably lower among the three youngest groups, and comparably higher in the three oldest groups.

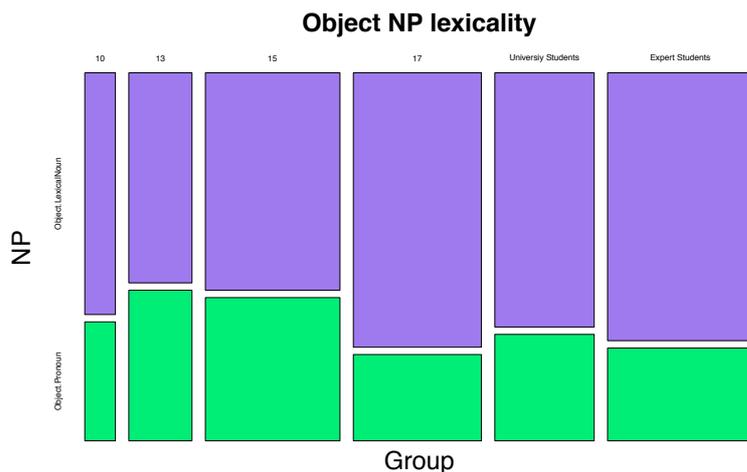


Figure 7. Proportion of lexical (purple) and pronominal (green) noun phrases in the Object constituent.

4.1.5 Other noun phrase lexicity – quantitative analysis

Finally, a significant group effect on the proportion of lexical noun phrases in Other function ($F(5, 90) = 4.43, p = 0.001$) was found. Post-hoc tests revealed significant differences between 10-year-olds and 17-year-olds ($p = 0.026$), between 10-year-olds and University Students ($p = 0.009$) and between 10-year-olds and Expert Students ($p = 0.014$). Furthermore, significant differences were found between 15-year-olds and 17-year-olds ($p = 0.015$), between 13-year-olds and University Students ($p = 0.029$), and between 13-year-olds and Expert Students ($p = 0.046$). Results are shown in Figure 8 and the pattern observed for lexical content in the Other function seems to correspond quite closely to the one for the Object function, even though the proportions of lexical NPs were consistently somewhat higher here.

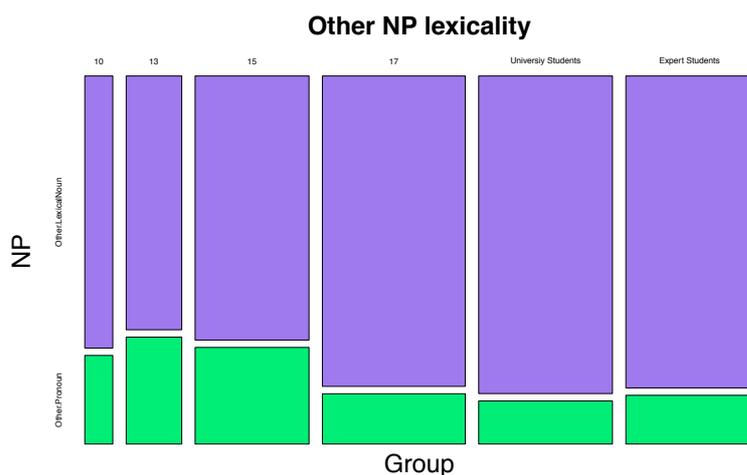


Figure 8. Proportion of lexical (purple) and pronominal (green) noun phrases in Other constituents.

4.2 Discussion of noun phrase lexicality results

As we have seen, the first hypothesis, predicting that the number of lexical noun phrases increase and the number of pronouns decrease with age and schooling was corroborated by the results reported above. Further, the result matched previous findings summarized in the Background section. What was not mentioned in the quantitative results was that three of the 10-year-old participants did not produce *any* lexical NP in the Subject function, and that two of them employed only two lexical NPs in this function, one of which produced two tokens of the same simple NP – *flickan* ('the girl'). Two 13-year-old participants also altogether lacked lexical NPs in the Subject function, and two of them produced only one. This phenomenon did not occur in any of the older four groups.

As we saw in section 2.2, Hunt (1965) attributes this change to the fact that younger writers chunk a corresponding informational content into a relatively greater number of clauses. With more clauses, they need to use pronouns to carry reference along from clause to clause, whereas a more trained writer would pack the same, or more, information into fewer but longer clauses. In example (1) we can see this pattern at work in a text by one of the 10-year-old subjects. We can also see a few other salient aspects of pronoun usage that will be discussed below. In five clauses (clause boundaries are marked '|'), the first writer uses seven pronouns, marked in bold (including the expletive subject *det* – 'it', or 'there').

(1) sedan var **det hon**| som tog pengarna| då måste **man** ju ge tillbaks pengarna| men **jag** vet inte | hur **man** ska få **dem** att göra **det** [wg18m]
'then it was she who took the money then you have to give the money back but I don't know how to make them to do that'

A mature writer, according to Hunt, would condense the information into fewer but longer clauses, with an increased number of modified lexical nouns as well as an increased number of modifiers to each noun. A contrasting sample of this salient pattern from an adult and substantially more experienced writer who conveys a great amount of information in two clauses, using no pronouns but several long, complex and coordinated lexical noun phrases is given in example (2) below, from one of the University Students:

(2) Grundläggande gäller alltså för samhället i stort att med större generella satsningar på den allmänna välfärden och med specifika satsningar inom skolektorn försöka få bukt med de problem| som existerar inom skolan. [wu19m]
'Basically and concerning society at large, it has to make more general investments in public welfare, and with specific initiatives within the school sector try to overcome the problems | that exist in schools'

Additionally, it can be observed that in example (1) above, there are several instances of unclear pronominal reference, another aspect of pronoun employment in the context of writing maturity (e.g. Perera 1984:156). The writer starts out referring to a character and an event in one of the short sequences in the elicitation movie, identifiable to him but potentially not known to the reader of the text, as *hon som tog pengarna* ('she who took the money').

(Naturally, the children may intuitively have addressed the investigator in their writing, a person who ought to be familiar with the characters in the film, thus overlooking the demands of the expository text type.) Reference subsequently changes to people (or children) in general in *man* (generic pronoun ‘one’), then a personal, first person stance is taken in *jag vet inte* (‘I don’t know’), returns to indefinite *man* and, finally, an unclear reference is made to unidentified people (or children) in *dom* (‘they’). Recognizing the demand for establishing common ground, and the needs of the reader, we find that both the given-ness and the accessibility of the referents (Chafe 1994), though self-evident to the *writer*, are severely hampered for the *reader*. For further examples, see (6) and (7) below, both produced by 10-year-old writers.

Whether unclear pronoun usage in example (1), or the identification via pronoun plus relative clause in the examples (3-7) can be attributed to an insufficient command of pronominal reference or vocabulary limitations, or an interaction of these factors, is open to interpretation. It is a frequent strategy among the 10-year-olds, to use a personal pronoun modified by a restrictive relative clause to identify a referent. This seems to be done in place of, for example, opting for a lexical noun that would be a more specific and succinct choice (examples 3-5 below, substituting the Swedish translations of ‘the bullies’, ‘bullying victim’ or ‘cheats’ in turn) or of a condensed construction using a pre-positioned adjective or an infinite construction (‘a bullying person’). Perera makes a comment on the high ratio of this structure in young children (about five), as a way to make up for a limited vocabulary (1984:146). In the instances cited here, we may speculate that the participants are familiar with the vocabulary per se, but with highly taxed working memory resources due to e.g. the structural demands of the less familiar expository genre, lexicon access is restrained.

- (3) många av dom som mobbar [wj04m]
‘many of those who bully’
- (4) en som är mobbad [wj05m]
‘one who is bullied’
- (5) de som fuskar på prov och sånt [wj10m]
‘those who cheat on tests and the like’
- (6) han som kom när de spelade basket [wg04m]
‘he who came when they played basket-ball’
- (7) hon som kom gående mot några tjejer [wg04m]
‘she who was walking towards a couple of girls’

This same construction, personal pronoun modified by restrictive relative clause, occurred at about the same ratio in all groups, 7-10% of all pronominal noun phrases, but is in fact slightly more frequent in texts written by the more experienced writers in the University Students and Expert Students groups. However, among these writers it was used in a distinct function, pointing to abstract or indefinite referents with certain characteristics, as in

- (8) den som blev mobbad i skolan [wu04f]
‘s/he who was bullied in school’

It is interesting at this point to remind ourselves of one of the more subtle characteristics of later language development, expressed so aptly by Scott: "Thus, the study of syntactic development in older children and adolescents requires sensitivity to the range of possible meanings encoded by any one structure" (1988:51-52). The two usages of the relative clause can be described in the same grammatical terminology, but below the surface, the function has changed from the concrete to the abstract.

The increase of lexicality with age, in this material, was especially salient in the Subject constituent, something that corroborated the fourth hypothesis. Pronominal noun phrases are "lighter" than lexical ones (Chafe 1994): used correctly they refer to known, or "given" referents, perceived so by writer and reader mutually, and they are, generally, easily retrievable – frequent, short, easy to spell and less structurally complex. The fact that the number of lexical noun phrases in the Subject function was so low compared to those in Object and Other functions supports the comprehension that the Subject constituent (in a preverbal position) is specifically vulnerable, and is more difficult to fill with lexical content (Chafe 1994, Ravid et al. 2002). This is consistent with the conventions of informational structure, where given information in the beginning of a sentence, as a rule, is expressed by a pronoun, and the new information, expressed by lexical nouns, is positioned at the end of the sentence. To my knowledge, the extent of the difference in lexicality between Subject, Object and Other positions has not been examined previously, and not in a developmental context.

4.3 Noun phrase expansion – length in number of words

In this section we test the second, fourth and fifth hypotheses. Consequently, the mean length of all lexical noun phrases was calculated for each group, in number of orthographic words, firstly for all NPs independent of syntactic function, secondly for the three constituent categories separately.

4.3.1 Noun phrase length – overall quantitative analysis

A highly significant effect of group on the length of lexical noun phrases in all clause constituents ($F(5, 4470) = 9.02$ $p = 0.000$) was found. Post-hoc tests showed that 10-year-olds produced significantly shorter noun phrases than University Students ($p = 0.001$), 13-year-olds shorter than 17-year-olds ($p = 0.042$), and than University Students ($p = 0.000$). Furthermore, significantly shorter NPs were produced by 15-year-olds than by University Students ($p = 0.000$), and by 17-year-olds than by University Students ($p = 0.001$). Finally, a negative significant difference was found between University Students and Expert Students ($p = 0.000$). Mean noun phrase length in number of words for each group is shown in the first column of Table 8. Average noun phrase length of the 13-year-olds ($M=2.27$) was found to be lower than that of the 10-year-olds ($M=2.44$), though this difference was not significant. Another result that stands out is that, on yet another measure, the Expert Students had cut back and produced significantly shorter lexical NPs ($M=2.54$) than the University Students ($M=3.13$), who produced the longest NPs. The Expert Students also produced below the mean of all groups ($M=2.69$).

Table 8. Mean lexical NP length in no of words per group, total and across clause constituents, minimum and maximum values within brackets.

	all constituents	Subject	Object	Other
10 years	2.44 (1-11)	2.92 (1-11)	2.21 (1-7)	2.33 (1-9)
13 years	2.27 (1-15)	1.89 (1-7)	2.45 (1-13)	2.36 (1-15)
15 years	2.53 (1-29)	2.42 (1-15)	2.84 (1-29)	2.37 (1-18)
17 years	2.74 (1-26)	2.49 (1-26)	3.19 (1-20)	2.61 (1-20)
University Students	3.13 (1-32)	2.52 (1-21)	3.90 (1-21)	3.15 (1-32)
Expert Students	2.54 (1-18)	2.14 (1-18)	2.88 (1-18)	2.53 (1-17)
total	2.69 (1-32)	2.38 (1-26)	3.05 (1-29)	2.65 (1-32)

4.3.2 Noun phrase length across clause constituent – quantitative analysis

Mean noun phrase lengths in number of words across clause position are displayed in Table 8 and in Figure 9. The longest Subject noun phrases, (M=2.92), were produced by the youngest age group, the 10-year-olds, and Subject NP length in the 13-year-olds was significantly lower, demonstrating the lowest mean of all groups (M=1.89).

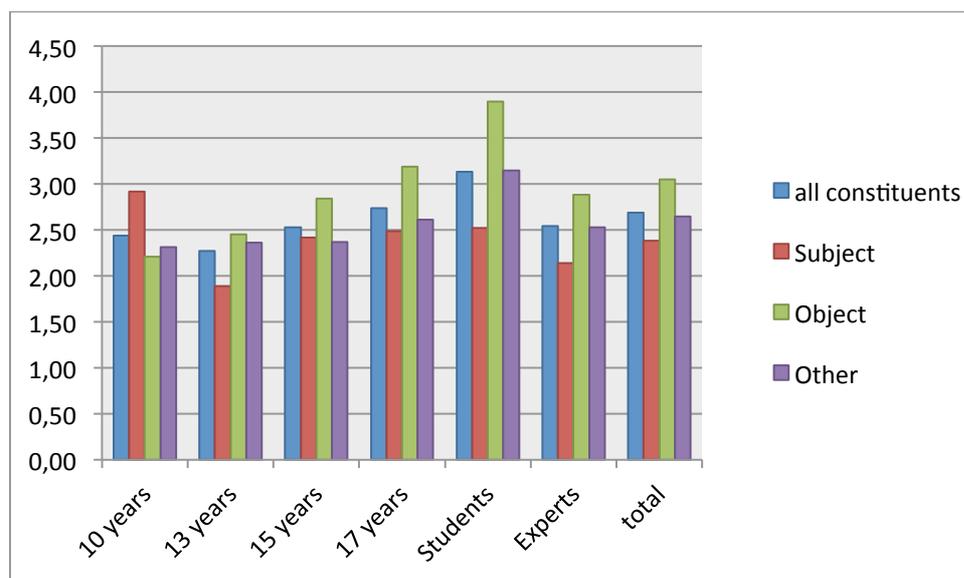


Figure 9. Mean length in number of words of lexical noun phrases, across group and clause constituent.

Significant effects of group on the length of lexical noun phrases in all syntactic functions were found, however less so for Subject ($F(5, 1188) = 2.25, p = 0.048$), than for Object ($F(5, 1265) = 6.25, p = 0.000$) and for Other ($F(5, 2004) = 5.53, p = 0.000$). Furthermore, significant negative differences for Subject constituent were found between 10-year-olds and 13-year-olds ($p = 0.011$) and, also negative, between 10-year-olds and Expert Students ($p = 0.028$). No further significant differences between groups were found for the Subject constituent.

For Object constituent the post-hoc analyses yielded significant differences between 10- and 17-year-olds, ($p = 0.016$), and between 10-year-olds and University Students, ($p = 0.000$). Additional significant differences were found between 13-year-olds and University Students ($p = 0.000$), and between 15-year-olds and University Students ($p = 0.001$). Finally, significant differences between 17-year-olds and University Students, ($p = 0.006$), and between University Students and Expert Students a negative difference ($p = 0.000$) were revealed.

For Other constituent significant differences were found between 10-year-olds and University students ($p = 0.005$), between 13-year-olds and University Students ($p = 0.008$), 15-year-olds and University Students ($p = 0.000$), 17-year-olds and University Students ($p = 0.001$), and finally, between University Students and Expert Students (negative $p = 0.000$).

4.4 Discussion of noun phrase length results

The second hypothesis, stating that noun phrase length in number of words increases with age, was mainly corroborated by the above findings for noun phrases overall. Thus, the results also mostly agreed with the conclusions from previous research as reported in the Background section. It is especially interesting to note that the results of this analysis echoed the ones of Ravid & Berman (2010) precisely: the 13-year-olds produced shorter noun phrases than the 10-year-olds. However, neither in that study nor in the present one do we learn what the actual phrases look like, i.e. whether certain types of modification are more favoured in certain groups. We return to the question of noun phrase complexity in the next Result subsection.

Further support for the second hypothesis was gained from the fact that University Students produced longer noun phrases than all other groups, for each and all of the clause constituents, with only one exception: the subject noun phrases employed by the 10-year-olds. We will return to those shortly. Additionally, the longer noun phrases produced by the University Students' group may be one of the explanations to why, as we may remember, the correlation between text length and number of noun phrases was dramatically lower for this group (section 3.2). One possibility is that a higher amount of information is conveyed in each noun phrase, thus decreasing the demand for a higher number of them. However, even though a low correlation is true for the group, we do not know what patterns occur on the individual level.

Moreover, aligned with the fifth hypothesis, the Expert Students produced, overall and in Object and Other functions, significantly shorter noun phrases than the University Students, the 17-year-olds *and* than the average of all groups, though not significantly so. Additionally, Subject NP length was found to be lower, though not significantly so. In all probability the higher lexical diversity - employing more precise and low-frequent words - played a part in this result, but did the Experts also employ, in general, more concise structures of modification? Did they produce adjectival, phrasal, prepositional attributes and compounds, rather than clausal attributes? Both Hunt (1965) and Loban (1976) advocate conciseness and economy as salient features of writing maturity. Taking reader-orientation into consideration,

expert writers may make an intuitive estimation of a limit to length, and to how much information can comfortably be digested within one single noun phrase. In this group too, there may be decisively individual styles, and/or the numbers may conceal a great variation within the category of lexical noun phrases, a rhythmical exchange between long and short. We will return to the question of the characteristics of the Expert Students' writing in the concluding discussion (section 5).

The fourth prediction, stating that development varies across the syntactic constituent, was also corroborated by the present results. Like in Einarsson's study, the Object constituent was longer than the Other constituent, and the Subject NP the shortest. This applied, generally, to all groups except the 10-year-olds, a somewhat surprising finding we will look into presently.

The youngest age group produced substantially longer lexical noun phrases in the Subject function than any other group, including the University Students (Table 8). The 10-year-olds also produced longer NPs in Subject than in Other and Object constituent, which is the reverse pattern to the other groups. Following an intuition, and at closer inspection, we found that a common NP structure produced by 10-year-olds, and most saliently in the Subject function, is a lexical noun in the definite form modified by a restrictive relative clause. This structure is used for the identification of a concrete referent in a concrete situation in the elicitation movie. We have already seen this pattern at work in the introductory text sample produced by a 10-year-old (Figure 1) – *dom två pojkar som bråkade* ('the two boys who were fighting') (wg07fCEW). Furthermore, the same structure, used with pronouns as heads, was seen in the investigation of lexicality. It is this construction, in this function, which is responsible for the relatively longer lexical noun phrases in the Subject function of the 10-year-olds. Out of eleven complex and/or more than five words long Subject NPs, and in this group, ten are constructions of that kind.

Comparing a second typical example:

- (9) flickan som kom för att sätta sig jämte två andra flickor [wg02f]
'the girl who came in order to sit down next to two other girls'

to Hunt's (1965) *Moby Dick* example (section 2.2), our 10-year-olds seem to have taken a developmental step. Replacing two independent clauses, in two different sentences, with one sentence containing a dependant clause modifier represents a somewhat more dense information packaging. (Just like in the previous examples with pronouns, however, reference does not respond to the needs of a reader, who, presumably, has not seen the film. To do full service as an introduction to the character for the reader, the head noun ought to have been in the indefinite form, and modified by an indefinite article, *en flicka som kom [...]* ('a girl who came [...]).) Furthermore, it is interesting to compare the relative clause-modifier construction to the recapitulatory pronoun, according to Perera (1986b:501) a frequent grammatical construction used in speech:

- (10) And the witches/. **they** 'all 'haunted in the houses/ [12 years].

Both constructions start with (re-) establishing the theme (*the girl, the witches*) and continue with the new, important information, thus conforming to the familiar, resource-saving given – new organisation of information. At the same time, employing the relative clause construction as a replacement for the recapitulatory pronoun is one way of showing that writing makes use of different constructions than speech. See also the discussion of left dislocation in speech in Strömquist, 1996:19-21. The remaining (one out of eleven) complex NP produced by a 10-year-old is a simple definite noun, modified by a prepositional phrase attribute, also identifying a character:

(11) pojken i klassrummet [wg02f]
'the boy in the class room'

This structure can be viewed as a next developmental step, further condensing the information.

The specific function of the attributive relative clause described above, is entirely absent in the texts by the 13-year-olds in Subject function, and apparently it disappears gradually in the older age groups. There are a few examples among the 15-year-olds (N=9), and the 17-year-olds (N=3), but the usage is far less frequent in those older groups. A new calculation, excluding relative clause-modified Subject noun phrases (not to dismiss them but to demonstrate their influence on the mean length of the relatively few lexical subject noun phrases), gave at hand a substantially lower result, M=1.90 replacing M=2.92. Among the older age groups, the impression is that the relative clause modifier with the concretely "identifying" function, is gradually replaced with prepositional phrase attributes, exemplified in the following phrases, all produced by 17-year-old writers:

(12) damen med mobiltelefonen [wh18m]
'the lady with the cell phone'

(13) ungdomarna på filmen [wh18m]
'the young people in the film'

or with an adjectival modifier in:

(14) en glasögonprydd tjej [wh19m]
'a girl wearing eye-glasses', literally: 'an eye-glass-adorned girl'

Lexical noun phrases with a relative clause modifier in the Subject constituent are less frequent among University Students and Expert Students but they do occur. However, the justification for relative clause modification produced by these writers is distinct:

(15) pojkar som inte förväntas kunna eller vilja utföra skolarbete [XA05e]
'boys who are not expected to be able or to want to do school work'

In parallel with the relative clause-modified pronouns in the section on lexicality, they are used by the more experienced writers to point at an abstract person or sub-group with certain characteristics or experiences.

4.5 Simple or complex lexical noun phrases – complexity

To find out to what extent the third, fourth and fifth hypothesis was corroborated by the present material, all complex lexical noun phrases, defined as in the Method section, 3.3.5 (compounds consisting of two lexical nouns, lexical nouns modified by a noun in the genitive, a prepositional phrase, an infinitival phrase, a restrictive relative clause or other clausal attribute) were identified and classified. The prediction was that with age and schooling, as attested by preceding research, noun phrases grow more complex, but that complexity decreases in expert writers, and that the developments are distinct for the different clause constituents. The raw numbers of complex lexical noun phrases were first calculated in CLAN (FREQ). Following that, the ratio of complex lexical NPs out of all lexical NPs were calculated for each group (number of complex lexical NPs/total number of lexical NPs), and then separately for Subject, Object and Other clause constituents.

4.5.1 Noun phrase complexity – quantitative analysis

A significant effect of group was found on the proportion of complex lexical noun phrases out of all lexical NPs irrespective of syntactic function ($F(5, 90) = 5.98, p = 0.000$). Post-hoc analyses revealed significant differences between 10-year-olds and the three older groups: 17-year-olds ($p = 0.004$), University Students ($p = 0.000$) and Expert Students ($p = 0.001$). Further significant differences were found between 13-year-olds and University Students ($p = 0.005$) and between 15-year-olds and University Students ($p = 0.037$). An overview of the overall proportion of complex noun phrases in the different groups is given in Table 9. A steady rise in complexity seems to take place for every consecutive group except for the Expert Students, who on yet another measure back down as compared to the peak number produced by the University Students.

Table 9. Mean percentage of complex noun phrases out of all lexical noun phrases across constituent and group.

	10 years	13 years	15 years	17 years	University Students	Expert Students	total
overall	21.8%	24.3 %	26.6%	32.1%	37,2%	34.1%	29.4%
Subject	14.6%	11.3%	19.9%	31.9%	31.1%	25.6%	22.4%
Object	22.4%	30.4%	31.9%	38.4%	47.3%	33.4%	34.0%
Other	17.9%	20.9%	25.9%	28.6%	37.6%	38.8%	28.3%

In Figure 10 it can also be seen that, to an even higher extent than the for the lexical noun phrases, the actual numbers of complex noun phrases produced by the 10-year-olds ($M=2.9$) constitute no more than a fraction of the numbers produced by 17-year-olds ($M=21.3$), University Students ($M=23.4$) and Expert Students ($M=21.8$).

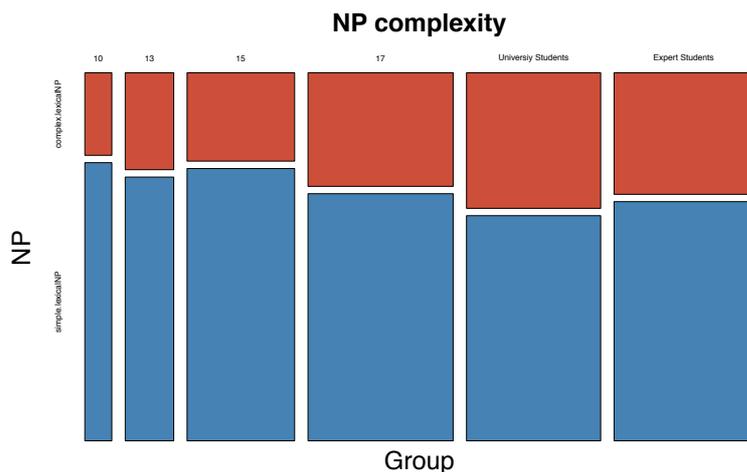


Figure 10. Proportion complex (red) and simple (blue) noun phrases out of all lexical noun phrases, by group.

4.5.2 Complexity across clause constituents – quantitative analysis

In Table 9 an overview over the mean proportion of complex noun phrases is given for all groups and syntactic functions. Contrary to the proportion of lexical noun phrases reported in section 4.1, all constituents account for a substantial increase in lexicality, and percentages are not as markedly different between clause constituents. Results for noun phrase complexity in the Subject, Object and Other constituents are displayed in Figure 11, Figure 12 and Figure 13 respectively.

For complexity in Subject function, a significant effect of group ($F(5, 90) = 4.75, p = 0.001$) was found. Post-hoc analyses yielded significant differences between 10-year-olds and 17-year-olds ($p = 0.002$), and between 10-year-olds and University Students ($p = 0.004$). Further significant differences were found between 13-year-olds and 17-year-olds ($p = 0.005$) and between 13-year-olds and University Students ($p = 0.007$). Finally, there was a significant difference between 15-year-olds and 17-year-olds ($p = 0.034$).

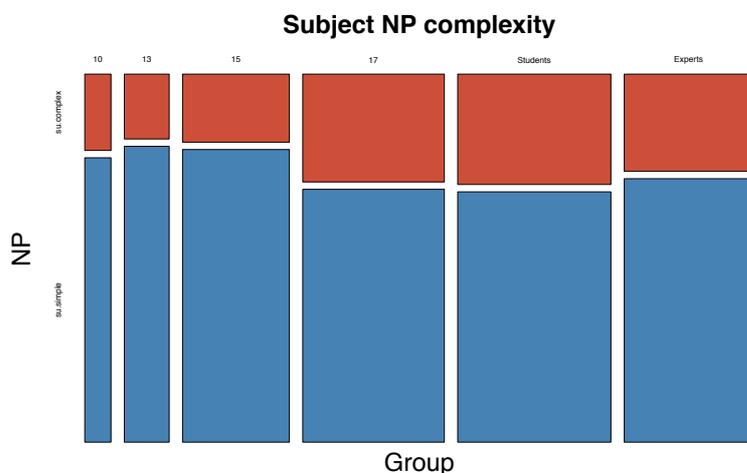


Figure 11. Proportion of complex (red) and simple (blue) lexical noun phrases as Subjects, by group.

For Objects additionally, analyses yielded a significant effect of group on the proportion of complex noun phrases ($F(5, 90) = 3.79, p = 0.004$). Significant differences were found, in post-hoc tests, between 10-year-olds and 17-year-olds ($p = 0.002$), and between 10-year-olds and University Students ($p = 0.004$). Finally a negative significant difference was found between University Students and Expert Students ($p = 0.024$).

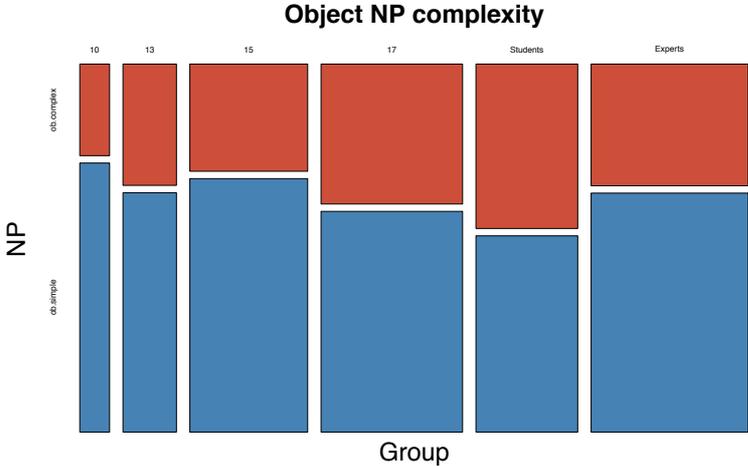


Figure 12. Proportion complex (red) and simple (blue) lexical noun phrases as Objects, by group.

Finally, the ANOVA showed a significant effect of group on the proportion of complex noun phrases in Other constituents ($F(5, 90) = 7.02, p = 0.000$). Post-hoc tests yielded significant differences between 10-year-olds on the one hand, and 17-year-olds ($p = 0.021$), University Students ($p = 0.000$), and Expert Students ($p = 0.000$) on the other. Furthermore, differences were shown to be significant between the group of 13-year-olds and University Students ($p = 0.005$), and finally, between 13-year-olds and Expert Students ($p = 0.002$).

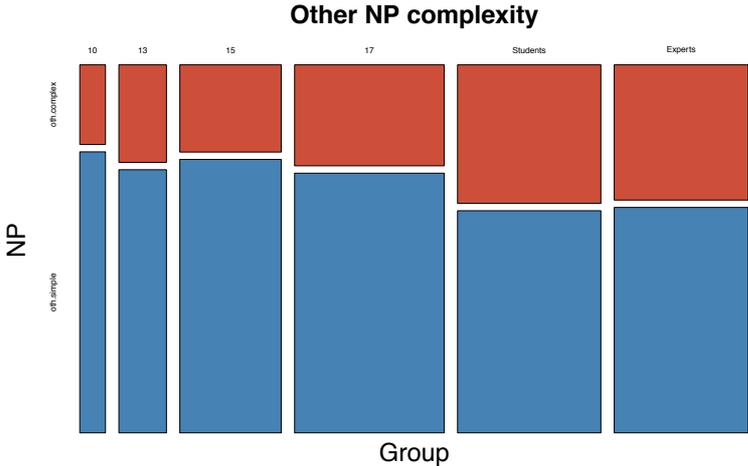


Figure 13. Proportion of complex (red) and simple (blue) lexical noun phrases as Other constituents, by group.

4.5.3 Discussion of noun phrase complexity results

Taken together, the results of the analyses of the present material supported all three hypotheses that concern noun phrase complexity. The percentage complex noun phrases increased with age and schooling for lexical noun phrases overall. It increased steadily from age group to age group up until the point where we observe the Expert Students taking a step back, producing a lower proportion of complex NPs than the University Student group. Additionally, the percentage complex noun phrases in the distinct clause constituent categories differed greatly, in the same pattern as that of lexicality and length: a lower percentage for the Subject category than for Object and Other. The results for subject complexity replicated those of earlier studies (Ravid et al. 2002). By this and several other accounts, and in line with “the light subject constraint” (Chafe 1994), the subject was shown to be the more “sensitive” of the constituents.

As could be expected, for the 10-year-old group, complexity overall was lower than in all other groups, and than the total mean. Additionally, this was true for Objects as well as for Other constituents. When it came to Subjects, it was higher for the 10-year-olds than for the 13-year-olds, though this difference was not significant. When we analysed noun phrase length, we saw that the frequent presence of the noun + relative clause-modifier construction in the noun phrase Subjects produced by 10-year-olds was to a great extent responsible for the fact that lexical subject noun phrases produced by this group were longer. There is reason to believe that the relative infrequency of this construction in the output by the 13-year-olds affected complexity results, too.

For the 17-year-olds, there was a significant rise in subject complexity compared to that of the preceding group, the 15-year-olds, and the ratio for 17-year-olds even exceeded that of the two subsequent adult groups, though not significantly so. As we expected, University Students showed high ratios of complexity in all categories, but especially for Object noun phrases. In the case of Objects, the negative difference to the Expert Students was significant. These results for complexity in University Students and 17-year-olds are interesting also in the context of Hunt’s (1965) and Loban’s (1976) views of a higher conciseness: fewer clauses and more complex phrases were interpreted as tokens of syntactic maturity.

We note that 13-year-olds produced proportionally less complex noun phrases in the Subject constituent than their younger peers, the 10-year-olds as subjects, and that Expert Students produced fewer complex noun phrases overall and in all constituents except Other, than the University Students. The reasons for these declines are probably different for the two groups. It lies near at hand to identify the strategy of the two youngest age groups as “knowledge-telling”, resulting in less concise and elaborate products. The greater complexity employed by the 10-year-olds can be attributed to their frequent usage of the less concise relative clause modification. The 13-year-olds seemed to have deserted relative clauses to a high extent, and moreover, they may not have moved on to the employment of prepositional phrases, as it appears, next in turn developmentally. We ask ourselves why the Expert Students produced a lower complexity in the Subject function than the University Students. We will return to that question in the concluding discussion to follow.

Furthermore, and also in the light of Scott's (1988) statement that most noun phrase expansion in writing takes place post-verbally, it is intriguing that the most significant and clear-cut increase and development of complexity (as of lexicality) in this work occurred in the Subject constituent. Concerning the length measure, as we remember, results were more mixed, depending on the frequent usage of the relative clause modification.

5 Concluding discussion

In this section, we shift the point of view somewhat and discuss the results going out from the two distinct noun phrase developments at issue in the present investigation. We remind ourselves that the developments discussed concern the expository text type. Results could very well turn out differently were we to investigate noun phrases in another genre. Most certainly, we have seen individual variation within the different groups, so what we discuss here are some of the general trends. First, we devote ourselves to the “typical” development occurring with age and schooling, and hereafter the expert development that ensues for writers training professionally. Finally, a few comments are made on directions for further research.

5.1 Age-related development and the noun phrase

As starting points for the quantitative analyses of age-related noun phrase development carried out within this work, four hypotheses were formulated on the basis of results from previous research in the area. Due to the paucity of investigations on Swedish, these hypotheses were mainly based on English data. Predictions were made that an increase would occur in lexicality, length and complexity of the noun phrase, with the age and schooling of the writers, and a variation in those developments depending on the syntactic function of the noun phrase (Subject, Object or Other). To a great extent, the results for the Swedish material in the present work corroborated the four hypotheses, but some interesting reservations were also made. In this section, we focus on a discussion of age-related development and highlight some interesting findings. We start out by noting that the timing of development in the respective aspects differs, and we go on to problematize the noun phrase complexity aspect. After that we discuss the fact that development is not straightforwardly linear, and conclude by pointing out that, even so, the University Students are – quantitatively and in the current measures – best in class.

5.1.1 Timing

Significant changes occurred with the group of 17-year-olds for most of the measures of noun phrase development. This applies to lexicality overall, as well as to lexicality in the Subject, Object and Other constituents separately. For overall and Subject lexicality, the step is more gradual in that the significant difference came about between the age of 10 and 17. For Object and Other lexicality, however, the significant difference occurred more abruptly between the 15-year-olds and the 17-year-olds. Additionally the rise in the percentage complex noun phrases for overall and for Object and Other constituents reached a significant level between the youngest age group and the 17-year-olds. For Subject constituent complexity, the significant change occurred between 15-year-olds and 17-year-olds. These results support the view that a developmental leap was taken by the 17-year-olds. They also give some support to the interpretation of the findings in the study by Gustafson (2012): The significantly higher percentage pause time spent in a phrase context during writing, indeed resulted in more lexical and complex noun phrases produced by the 17-year-olds as a group. However, not until the correlations between longer pause time and more lexical and complex noun phrases

are tested on an individual level can we conclude from this investigation that product was found to reflect process.

The length measure, in its turn, peaked with the University Students. Thus, the significant changes for noun phrase length in all constituents taken together, and in Object and Other constituents separately, took place between the group of 17-year-olds and the group of University Students. For the Object category there was an additional significant change between 10-year-olds and 17-year-olds. For Subject length the results were less clear-cut, due to the frequent usage of the relative clause modifier by the 10-year-olds, and, as we remind ourselves, they showed a negative significant difference between 10-year-olds and 13-year-olds, and additionally between 10-year-olds and Expert Students. Making Hunt's (1965) and Loban's "conciseness" and "economy" our beacons of writing maturity, the higher noun phrase length found in the group of University Students group may at first seem contradictory. However, long noun phrases may be the outcome of co-ordinations and combinations of noun phrases and of modifiers that in themselves may be concise.

One of the findings of the present work was that, in the youngest age group, the relative clause modifier was frequently used and affected length and complexity results substantially, not least in the Subject constituent. This is especially interesting since previous researchers on English (Hunt 1965, Loban 1976, Perera 1984), frequently point to the relative clause as a structure indicative of increasing linguistic writing maturity at the same time as a higher conciseness – fewer clauses and more complex noun phrases – are prized as tokens of the same maturity. A tentative conclusion is that the partiality in Swedish for the relative clause affects the development towards a greater conciseness (employing one main clause and one subordinate clause within the same sentence instead of two main clauses). And at the same time, constructions like the recapitulatory pronoun, characteristic of speech, are avoided. Presumably the condensation into pre-posed adjectives or post-posed infinite and prepositional constructions represents a step two-strategy. This fact urges us to return to the question of what complexity is.

5.1.2 The complexity aspect

Out of the four aspects investigated in the current analysis, and as was stated in the Method section, complexity is certainly the most problematic one to define. The concept of complexity used in the present work was based on research emphasizing that post-modification tends to grow the most in later development and therefore, preliminarily, could be viewed as the most complex. Our results, however, showed that, for one example, the post-positioned relative clause-modifier was used by the youngest age group at least as frequently as by the two adult groups, and more frequently than the intervening three adolescent groups. The conclusion now must be that complexity is more complex (!) than that and, based on the results of the analyses of noun phrase length measures, a more detailed analysis of the different types of modifiers must be made. A scan for the long noun phrases in the University Students' group yielded samples like:

- (16) ett djur med samma behov och känslor som en jagande flock eller en jagad hjord [ws03m]
'an animal with the same needs and feelings as a hunting flock or a hunted herd'

It seems clear that to find out more about complexity, we need to analyse more closely the relative frequency of the different complex constructions – compounds, genitive constructions, prepositional phrases, infinitive phrases, relative clausal and other clausal modifications – and to when they occur during development. Secondly, the usage of coordinated noun phrases and modifiers must be investigated, as well as the combination of different attributes and the number of attributes to each noun phrase. Incidentally, the material was coded for coordinated noun phrases and modification within noun phrases, but due to the limitations of this work, these analyses had to be put aside.

5.1.3 Development is not straightforward – the case of the 13-year-olds

One reservation concerning the support for the four predictions on noun phrase development is that, in the current aspects, lexicality, length and complexity did not grow in a straightforwardly linear way. This was clearly illustrated by the fact that 13-year-olds produced less lexical noun phrases in all measures, as well as shorter noun phrases overall and as Subjects, and less complex noun phrases as Subjects than the 10-year-olds, though only the difference in Subject NP length reached a significant level. And in addition to the results of the present work, earlier research (Johansson, 2009) reports that, for the same participants, and in the written expository texts, lexical density is significantly lower. We ask ourselves what the participants of this age group are preoccupied by, that does not compete for attention in the 10-year-olds. We remind ourselves that 13-year-olds produced texts that contained nearly the double amount of words compared to 10-year-olds, though variation was large within both groups. Furthermore, they may have been working with the thematic structure of the text, as described by Annette Karmiloff-Smith (1981). Several things point to the possibility that they are, to a greater extent, preoccupied with following text type conventions. We have seen a substantial decrease in the reading comprehension of fourth-graders when they are abruptly exposed to the expository (Snyder & Caccamise, 2010). The new demands of the unfamiliar genre must surely affect production as well. The 13-year-olds may be fighting with the distinct structure of the expository. Moreover, when we look further into the details of pronoun usage we see that, below the radar, an additional development is taking place.

The group of 13-year-olds, thus, is the one with the highest proportion of pronouns, both taken together, and for each of the syntactic constituents separately. We know from Hunt (1965) that one trend that would contribute to a lower pronominal ratio of the older writers is that of condensation – 13-year-olds (and 10-year-olds) possibly use finite clauses (*du ska inte stjäla*: 'you shall not steal') where older writers use infinite constructions (*att stjäla är dumt*: 'stealing is bad'). But we ask ourselves what the actual proportions of different pronouns are, within that category.

Going out from an intuitive impression, a quick calculation of occurrences of the indefinite pronoun *man* (English ‘one’, or ‘you’ in an indefinite sense) showed that for the 13-year-olds, nearly 27% of the pronouns category consisted of the single word *man*. This was substantially more than for both the one younger (17%), and for the four older groups (11%, 17%, 18% and 13% respectively). It could be the case that a number of the 13-year-old participants have recently discovered the application of the indefinite pronoun *man*, that this constitutes part of their understanding of the new genre, and, as is often the case with newly acquired competencies, tend to over-use it.

Comparing the frequency of *man* (‘one’, or ‘you’) with counts for the pronoun *det* (English ‘it’, or ‘there’ as in ‘there are ...’), a pronoun both frequently used as an expletive subject and as a personal pronoun, we found that 13-year-olds seem to have reached an adult level – 10-year-olds use it only in 17% of all pronominal NP occurrences, but the ratio is around 24-26% for all of the remaining groups. Consequently, we would want to know what the figures for the distinct groups look like if we separate the expletive subject from the usage of personal pronouns.

There are other patterns that hint in the direction of 13-year-olds being preoccupied with the gradual acquisition of genre conventions. One example is that the pattern of referring to characters in the elicitation movie that the 10-year-olds frequently use is not occurring to the same degree at all in the 13-year-olds’ group. This, together with the increased usage of the indefinite pronoun *man*, may imply that 13-year-olds are on their way to generalising on the topic of “problems between people”, in contrast to the 10-year-olds re-narrating the specific events of specific characters.

5.1.4 University students best in class

Even if age-related development, as was exemplified above, is not simply a linear movement from less to more, the adult University Students definitely stood out as best in class on most counts. They did not produce significantly less or lower than any other group on any of the measures. In the study, they generally produced the longest (though 10-year-olds produce longer Subject noun phrases), the most lexical (though Experts and 17-year-olds produce a higher percentage of lexical Objects) and the most complex noun phrases (though Experts produce more complex NPs in Other functions), and always above the overall means.

However, when the correlation between number of words in text and number of noun phrases was tested, and all other groups show a high correlation, 0.9-0.98, University Students show a saliently low correlation, 0.14 (Table 4). This fact, together with the higher percentage long and complex noun phrases imply that fewer noun phrases carry a heavier informational load, containing more modifying words within the NP, at least for a substantial number of the participants. Example (16) shows how this can be carried out:

(17) möjligheten att släppa det instinktiva handlandet och själva avgöra hur vi ska mottaga nya medlemmar i vår flock [ws03m]

'the possibility to let go of instinctive action and decide for ourselves how to receive new members of our flock'

Without an additional adult group, the Expert Students, we could be satisfied with concluding that longer, more lexical and more complex noun phrases are at the end of the developmental scale – more is better.

5.2 Expert writing and the noun phrase

Our hypothesis number five concerning noun phrase development in expert writing predicted a decrease in noun phrase length and complexity. These predictions were made on the assumption that one aspect of reader-orientation is a simpler syntax. Generally speaking, the hypotheses were supported by the results of this investigation. Regarding overall noun phrase length a negative significant difference between University Students and Expert Students was found, and likewise in Object and Other functions taken separately. All numbers on length were lower than those of the University Students, *and* than the average of all participants. For complexity only the difference in Objects reached a significant level but apart from that, Expert Students had a lower proportion on all counts except in Other constituents. Incidentally, the results for lexicality revealed no significant differences between University and Expert Students, and tended to look much the same, possibly with the exception of Object noun phrases that were more lexical for the Experts, though a significant level was not reached.

We may assume that the Experts are, generally, capable of producing as long and as complex noun phrases as the University Students. We also assume that they are capable of making more conscious choices, of being designers. As a consequence of their experience and training, Experts have a more varied choice, and they do choose (as a group) to use simpler and shorter noun phrases. This can be due to stylistic considerations: the rhythm or flow of the text. One aspect of Kellogg's (2008) knowledge-crafting may well be rhythm. Another component may consist of a competent and dynamic variation. It may be that long and complex noun phrases do exist, and even to a high extent, in the Experts' texts, but that they more often take turns with shorter and simpler ones. Furthermore, an expert may have a more or less intuitive competence to make the text smoothly digestible for the receiver, and at the same time appreciating the need for variation to keep the reader attentive. Expert writers may be intrinsically aware of a limit to the amount of information comfortably processed within a single noun phrase. Additionally, creative writers are good at telling stories – they may be more prone to give examples, or illustrate facts and arguments with narrative passages, something which results in expositories that are, noun phrase-wise, less expository-like.

Questions of great interest for theories of the development of writing and of writing expertise are raised by the fact that expert noun phrase design is not simply the equivalent of more complex, or longer. On the contrary, on most of our measures, the Expert Students declined compared to the University Students. On other measures, the increase halted, something that also applies to lexicality. We know that individual variation is great, in age and experience, within this group. Nevertheless, we have obtained a result showing a distinct trend, a trend

that is compatible with an appreciation that reader-orientation entails shorter, and less complex noun phrases though at the same time, with the help of an extensive vocabulary, as concise. Not until a more detailed analysis has been carried out, however, will it be possible to pinpoint the characteristics of the noun phrases employed by the Expert Students, and what motivates the lower numbers in length and complexity.

5.3 Method discussion

The method used in the present analysis can be described as mainly quantitative, data-driven and descriptive. All methods have shortcomings and limitations that cannot be avoided.. Quantities and frequencies are important components in the study of language. As we have seen, later syntactic development, at least in part, consists of low-frequency structures becoming more frequent. The present thesis used both quantitative and qualitative components. A lot of time and effort was spent designing the coding schedule. Subsequently, each and every noun phrase was identified, evaluated and categorized by the investigator. The quantitative analyses, moreover, were to a great extent followed up with deep dives into the material with examples, and the observation and discussion of patterns and tendencies below the surface, and within each aspect of noun phrase development.

I have accounted for some questions concerning the definition of complexity used in the current work (5.1.2). Potentially, it could have been more rewarding to include coordinated noun phrases and attributes, and to take into account the combination of different attributes and this could have revealed another pattern.

Because I was interested in covering as much as possible of the developments, the material in this study derives from corpora that are not 100% comparable. It should also be mentioned that the texts that constitute the material were written in a lab setting, with a time constraint, and with no follow-up, factors, we may think makes the results less ecologically valid. The gain, on the other hand, is a large number of noun phrase samples for analysis. It would have been impossible, needless to say, to collect an equal amount of data within the limitations of the current task. A reasonable option could be to compare the results with national examinations in school and papers written at college. However, that would result in less control over instructions given, and in texts from different genres and with different topics. Experiments including elicitations of noun phrases, or a survey, interviews or self-reported tasks would arguably have given another material altogether, and would not have been as suited to test my hypotheses, or have had the same potential for generalizability. However, my basic approach is that any kind of study benefits from a complementary view.

A complex matter is the question of the respective roles, in language development, of educational instruction and cognitive evolution. What influence does instruction have when it comes down to it? I have, to a large extent, evaded this question in the study at hand, but this does not mean that I am not aware of it, and the fact that it is something that needs to be addressed.

5.4 Concluding remarks, further questions and directions for further research

As with most research we are left with more questions than answers. In this work we have seen some general tendencies, but also great variation. Many of the remaining questions concern a greater depth of analysis. We want to investigate individual variation within the groups in more detail. Can for example distinct “writer types” concerning noun phrase usage be found, especially among the adult writer groups? Do expert writers show the same range of constructions as the university students, even though their numbers are lower? Moreover, we have stated the need for a breakdown of the pronoun category, and the need to analyse the development of the different types of modification separately. What is the developmental timetable for compounding, for adjectival, prepositional, infinitive, prepositional and clausal modifiers? What implications will the results from analyses of that kind have for the concept of a complex noun phrase? Another question of great interest in this context is to what extent different types of attributes within the same noun phrase are combined, and to what extent noun phrases and modifiers within noun phrases are coordinated, and once more, what impact will the patterns found have on the notion of complexity? Further, what will be the results in lexicality, length and complexity if a separate analysis for Subject noun phrases in a position after the finite verb, and Object and Other noun phrases in the first position are compared to the overall figures for each constituent?

Furthermore, we set out to investigate two assumedly distinct developments: one “typical” age-related development and one expert development, and indeed, we found distinctions between the two. Still, we need to ask ourselves to what degree the “typical” development is typical. The pupils in the schools represented in the material were to a high extent expected to continue to higher education, and the University Students are adults who have written at least one BA thesis. Assuming the more than one adult standard of Scott (1988) we need to compare not only expert students with typical students but also both of them to adults who left school at 17 or earlier.

Another, potentially rewarding, topic to follow out is certainly the detailed comparison of noun phrases produced by University Students and those produced by Expert Students. Do the two groups employ as concise structures of modification: adjectival, phrasal, prepositional attributes and compounds, rather than clausal attributes? It would be of considerable interest to compare the effectiveness of information-packaging between these groups, too, possibly quantified as how many different attributes are attached to one nominal head in the respective groups.

In the current work we focussed on the product: the actual noun phrases in the final text. We have seen or hypothesized a number of diverse factors affecting the development of noun phrase design in an expository text. Additional factors are to be found in yet other cognitively motivated models. McCutchen (1996) emphasizes the role of working memory restrictions in many of the choices made during written language production. Writing is a complex activity, and has been described as a competition for cognitive resources by a number of component

processes. These processes consist of everything from spelling and word retrieval to global text planning. The development of writing expertise, according to Kellogg (2008) and McCutchen (2000), leans on experience and training which automatizes a successive number of those processes, economizing on working memory. Consequences of a shortage of resources relevant to the current topic could be when young and inexperienced writers use clauses in place of phrases, pronouns instead of lexical nouns, or simple lexical nouns instead of concise and complex ones. Using a research design including relevant working-memory data, correlations with noun phrase design can be explored.

Relatedly, a question of special interest is what the process behind concise information packaging looks like. Can the ability to perform concise information packaging in noun phrases be automatized or does it come as a result of extensive editing like the one characteristic of Expert Students (Gustafson et al., 2014). It has been shown that the efficient usage of revision is a competence generally hard to learn for children (Graves, 1979).

In this context, we make a short detour, demonstrating what may be revealed by data from the recorded writing processes. Even though the finished text product is in focus of the present work, the data collection procedure in all cases included keystroke logging in the software ScriptLog (Frid et al., 2012). This procedure makes it possible to analyse all steps of text production, and for example look at all possible changes and revisions leading up to the final noun phrase wording in the final text of each participant.

Specific to the question of edition, or non-edition, of the efficient noun phrase, we take a look at an excerpt from a linear text recording, containing all keyboard events during the writing process, in this case apart from the characters written backward deletions <BACKSPACE + how far> and pauses <duration in seconds>. The content of the noun phrase in the final text is effectively packaged as a pre-posed adjective attribute: *oförtjänta fördelar* ('undeserved advantages'). However, examining the linear text, we find that, in the first version, it consisted of a post-posed relative clause, *fördelar som är oförtjänta* ('advantages that are undeserved'):

(18) <BACKSPACE1>t skapar fördelar <2.871> som är fo<BACKSPACE1>
<BACKSPACE1>ofrä<BACKSPACE1> <BACKSPACE1>örtjänta.
'creates advantages that are undeserved' [XA07e]

This is an example of how informational content, as above-mentioned, can be packaged into a more dense form, and in this case through later revision.

Thus, in the recorded writing process, the context of example (18), at least, it is in fact possible to see that the dense noun phrase, containing one adjectival attribute, of the final product, the text, is in fact not produced immediately and automatically but starts out as a lexical noun modified by a longer relative clause attribute and is honed during a substantial phase of final editing, at the end of the whole writing process. The relation of the process to the product is indeed worth looking further into, though only one of the many fascinating aspects of noun phrase development that we have encountered in the current study.

References

- Bereiter, Carl and Scardamalia, Marlene (1987). *The Psychology of Written Composition*. Lawrence Erlbaum Associates, Hillsdale, NJ.
- Berman, Ruth and Bracha Nir (2010). The Language of Expository Discourse across Adolescence. In: Nippold and Scott (eds): *Expository Discourse in Children, Adolescents, and Adults*. New York: Psychology Press.
- Berman, Ruth and Verhoeven, Ludo (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language and Literacy*, 5(1), 1–44.
- Biber, Douglas (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Bohnacker, Ute (2003). Nominal phrases. In: Josefsson, G., Platzack, C. and Håkansson, G.: *The Acquisition of Swedish Grammar (195-260)*. Amsterdam: John Benjamins.
- Chafe, Wallace (1994). *Discourse, Consciousness, and Time*. Chicago and London: The University of Chicago Press.
- Delsing, Lars-Olof (1993). *The Internal Structure of Noun Phrases in the Scandinavian Languages: A Comparative Study*. Lund: Department of Scandinavian Languages, University of Lund.
- Einarsson, Jan (1978). *Talad och skriven svenska: sociolingvistiska studier*. Akademisk avhandling. Lund: Walter Ekstrands bokförlag.
- Fawcett, R. and Perkins, M. (1980). *Child language transcripts 6-12, Volumes 1-4*. Pontypridd, Mid Glamorgan, Wales: Department of Behavioural and Communicative Studies, Polytechnic of Wales.
- Fayol, Michel, Largy, P., & Lemaire, P (1994). When cognitive overload enhances subject – verb agreement errors. *Quarterly Journal of Experimental Psychology*, 47A, 437–464.
- Fox, John & Weisberg, Sanford (2011). *An R companion to applied regression*. Thousand Oaks, California: SAGE Publications.
- Frid, Johan, Johansson, Victoria, Johansson, Roger, Wengelin, Åsa & Johansson, Mikael. (2014). *Developing a keystroke logging program into a writing experiment*

- environment*. Poster presented at Writing Across Borders, WRAB, Paris 19–22 Feb 2014.
- Graves, D. H. (1979). RESEARCH UPDATE: What Children Show Us about Revision. *Language Arts*, 56(3), 312–319. Retrieved from <http://www.jstor.org/stable/41404804>
- Gustafson, Pia (2012). Written Language Development in Adolescents: Pause patterns and syntax in the writing process. Lund: Lunds universitet.
- Gustafson, Pia, Johansson, Victoria, Johansson, Roger, Wengelin, Åsa & Frid, Johan (2014). Reader-awareness in adult writers. An investigation of university students at a Creative Writing-program. Poster presented at the SIG-Writing Conference, Amsterdam, August 2014.
- Halliday, Michael A.K. (1985). *Spoken and Written Language*. Specialised curriculum: Language & Learning. Deakin University Press, Victoria.
- Harpin, W. (1976). The Second ‘R’. Writing Development in the Junior School. London: Allen and Unwin.
- Hultman, Tor (1987). Objekt som objekt. In Teleman, Ulf, editor: *Grammatik på villovägar*, 39-55. Stockholm: Svenska språknämnden och Liber.
- Hunt, Kellogg W. (1965). *Grammatical structures written at three grade levels* (Research Report No.3). Champaign, IL: National Council of Teachers of English.
- Hunt, Kellogg W. (1970). Syntactic maturity in school children and adults. *Society for Research in Child Development Monographs*, No. 134, 35, No. 1.
- Jackendoff, Ray (1977). X-bar-syntax: A study of phrase structure. Linguistic Inquiry Monograph 2. Cambridge, MA: MIT Press.
- Johansson, Roger, Johansson, Victoria, Wengelin, Åsa, and Holmqvist, Kenneth (2008). Dynamics of Perception and Production in Text Writing: Four Different Groups of Writers, volume 53 of Working paper in Linguistics, Lund University, pages 43–59. Centre For Language and Literature, Lund University, Lund.
- Johansson, Victoria (2009). *Developmental Aspects of Text Production in Writing and Speech*. Lund: Department of Linguistics and Phonetics, Lund University.
- Julien, Marit (2005). *Nominal phrases from a Scandinavian Perspective*. Amsterdam: Benjamin.

- Karmiloff-Smith, Annette (1981) The grammatical marking of thematic structure in the development of language production. In Deutsch, W. (ed), *The child's conception of language*, pp121-147. Academic Press, London.
- Kellogg, Ronald T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of writing research*, 1(1), 1-26.
- Loban, Walter D. (1963). *The language of elementary school children*. Champaign, Illinois: National Council of Teachers of English.
- Loban, Walter D. (1976). *Language Development. Kindergarten through Grade Twelve*. (Research Report No. 18). Champaign, IL: National Council of Teachers of English.
- Lundin, Katarina (2014). *Tala om språk: Grammatik för lärarstudier*. Lund: Studentlitteratur.
- MacWhinney, Brian (2015a). The CHILDES Project. Tools for Analyzing Talk – Electronic Edition.
- MacWhinney, Brian (2015b). The CHILDES Project. Tools for Analyzing Talk – Electronic Edition. Volume 2: Transcription Format and Programs. Part 2: The CLAN Programs.
- McCutchen, Deborah (1996). A Capacity Theory of Writing: Working Memory in Composition. *Educational Psychology Review*, Vol. 8, No. 3.
- McCutchen, Deborah (2000). Knowledge, Processing, and Working memory: Implications for a Theory of Writing. *Educational Psychologist*, 35:1, 13-23.
- McCutchen, Deborah (2006). Cognitive Factors in the Development of Children's Writing. In MacArthur, Charles A., Graham, Steve & Fitzgerald, Jill, editors, *Handbook of Writing Research*. Guilford, New York.
- McCutchen, Deborah (2011). From Novice to Expert: Implications of Language Skills and Writing- Relevant Knowledge for Memory during the Development of Writing Skill. *Journal of Writing Research*, 3(1), 51-68.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information'. *Psychological Review* 63 (2): 81–97.
- Myhill, Debra (2009). Becoming A Designer: Trajectories of Linguistic Development. In: Beard, Roger; Myhill, Debra; Riley, Jeni and Nystrand, Martin (eds.), *The SAGE Handbook of Writing Development*. London: SAGE Publications, 402–414.

- McNeill, David (1966). The creation of language by children. In: Lyons, J. and Wales R.J. (eds.) *Psycholinguistics Papers: The proceedings of the 1966 Edinburgh Conference*. Edinburgh: Edinburgh University Press.
- Nettelbladt, Ulrika and Salameh, Eva-Kristina (2007). *Språkutveckling och språkstörning hos barn. Del 1: Fonologi, grammatik och lexikon*. Lund: Studentlitteratur.
- O'Donnell, Roy C, Griffin, William .J. and Norris, Raymond C. (1967). *Syntax of Kindergarten and Elementary School Children: A Transformational Analysis*. Champaign, Illinois: National Council of Teachers of English, Research Report No 8.
- Perera, Katharine (1984). *Children's Reading and Writing: Analysing Classroom Language*. Blackwell, Oxford.
- Perera, Katharine (1986a). Grammatical Differentiation Between Speech and Writing in Children Aged 8 to 12. In: Wilkinson, Andrew, editor, *The Writing of Writing*. Open University Press, Milton Keynes.
- Perera, Katharine (1986b). Language acquisition and writing. In Fletcher, Paul and Garman, Michael, eds., *Language Acquisition*. Cambridge: Cambridge University Press.
- Platzack, Christer (2000): A Complement-of-N⁰ Account of Restrictive and Non-Restrictive Relatives. The case of Swedish. In: Alexiadou, A. et al. (eds.): *The syntax of relative clauses*, 265–308. Amsterdam: John Benjamins.
- Platzack, Christer (2011). *Den fantastiska grammatiken*. Norstedts, Stockholm, Sweden.
- R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Ravid, Dorit, van Hell, Janet G., Rosado, Elisa, & Zamora, Anita. (2002). Subject NP patterning in the development of text production: Speech and writing. *Written Language and Literacy*, 5, 69–94.
- Ravid, Dorit & Berman, Ruth A. (2010). Developing Noun Phrase Complexity at School Age: A Text-Embedded Cross-Linguistic Analysis. *First Language*, 2010 30:3.
- Ravid, Dorit, Dromi, Esther & Kotler, Pazit (2010). In: Nippold and Scott (eds): *Expository Discourse in Children, Adolescents, and Adults*. New York: Psychology Press.
- Roeser, Jens, Torrance, Mark & Baguley, Thomas (2015). Syntactic planning and lexical access in sentence production.
- Scott, Cheryl. M. (1988). Spoken and written syntax. In Nippold, M. A., editor, *Later*

Language Development. Ages nine through nineteen, pages 49–95. Pro Ed, Austin, Texas.

- Slobin, Dan I. (1971). *Psycholinguistics*. Glenview, Illinois: Scott, Foresman and Co.
- Snyder, Lynn & Caccamise, Donna (2010). Comprehension Processes for Expository Text: Building Meaning and Making Sense. In: Nippold and Scott (eds): *Expository Discourse in Children, Adolescents, and Adults*. New York: Psychology Press.
- Strömquist, Sven, Holmquist, Kenneth, Johansson, Victoria, Karlsson, H., & Wengelin, Åsa (2006). What keystroke logging can reveal about writing. In G. Rijlaarsdam (Series Ed.) and K. P. H. Sullivan, & E. Lindgren. (Vol. Eds.), *Studies in Writing, Vol. 18, Computer Keystroke Logging: Methods and Applications*, (s. 45–71).
- Strömquist, Sven (1996). Discourse flow and linguistic information structuring: Explorations in speech and writing. Göteborg: *Gothenburg papers in theoretical linguistics 78*.
- Strömquist, Sven (2009). *Språkets öga*. Lund: Studentlitteratur.
- Strömquist, Sven & Verhoeven, Ludo (2004). *Relating events in narrative – typological and contextual perspectives*. Mahwah, New Jersey: Lawrence Erlbaum Associate Publishers.
- Strömquist, Sven, Wagner, Åse Kari H & Uppstad, Per-Henning (2010). *Den flerspråkiga människan: en bok om skriftspråklärande*. Studentlitteratur AB, Lund.
- Teleman, Ulf, Hellberg, Staffan & Andersson, Erik (1999). *Svenska Akademiens Grammatik: 3. Fraser*. Stockholm: Svenska Akademien.
- Westman, Margareta (1974). *Bruksprosa*. Lund: LiberLäromedel.
- Wolfe-Quintero, Kate, Inagaki, Shunji & Kim, Hae-Young (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy & Complexity*. Mānoa: Second Language Teaching & Curriculum Center: Technical Report No. 17.