

**Subject NP patterning in the development of text production:
A crosslinguistic study of Dutch, English, Hebrew, and Spanish¹**

Dorit Ravid*, Janet G. van Hell, Elisa Rosado*** and Anita Zamora******

Ravid, D., J. van Hell, E. Rosado & A. Zamora. Subject NP patterning in the development of text production: Speech and writing. *Written Language and Literacy*, 5, 69-94, 2002.

Abstract

This paper examines how choice of subject NP types and structures changes in the development of text construction, and to what extent developmental patterns vary across languages in personal-experience narratives compared with expository texts produced in speech and in writing. The population for this study consisted of 80 participants, 40 gradeschool children and 40 university-level adults, with 20 participants in each of four languages: Dutch, Hebrew, English, and Spanish. The database for each language-specific analysis consisted of 40 gradeschool texts and 40 adult texts, half spoken and half written texts, half narratives and half expository texts in each group, altogether 320 texts. All subject NP slots in each text were counted and classified by category of realization (zero, pronoun, or lexical), by pronoun type (personal versus impersonal), and by lexical complexity (terminal NPs governing a single lexical noun versus nonterminal NPs governing more than one lexical noun). In general, the written expositives of adults are the preferred site for lexical subjects and for non-terminal subjects. Among both children and adults, narratives contain more personal subject pronouns, and expository texts contain more impersonal pronouns. Several crosslinguistic differences emerged, mainly between Spanish and the other three languages, reflecting differences in the syntactic, inflectional, and pronominal patternings of the target languages.

*School of Education and the Dept. of Communications Disorders, Tel Aviv University, Israel

**Faculty of Social Sciences, University of Nijmegen, The Netherlands

*** Institut de Ciències de l'Educació, Universitat de Barcelona, Spain

**** Department of Psychology, San Diego State University, USA

¹ xx small print We gratefully acknowledge the major conceptual contribution of Ruth Berman in developing the original NP templates. We also thank Melina Aparici, Sarah Kriz, Moniek van Oosterhout, Judy Reilly, Merit Sternau, and Marjan Tak for their invaluable assistance in developing the NP templates, in coding and analyzing the texts, and in discussing syntactic problems cross-linguistically.

1. INTRODUCTION

This study compares the distribution and internal architecture of subject noun phrases in spoken and written texts produced by gradeschoolers and adults in four languages. It examines how subject NPs change in structure and content across development and how these developmental patterns vary across languages in conveying the thematic content typical of personal-experience narratives compared with expository discussions, both spoken and written.

In the course of text construction, linguistic forms and categories interact with different discourse properties. Two such dimensions relevant to the present analysis are *modality* of production – the constraints imposed by online processes of producing speech versus writing -- and *genre* – the choice of text type, as defined by its thematic content and communicative function in socio-cultural settings (Ravid & Tolchinsky, xx in press). The nature and internal structure of linguistic elements in spoken and written texts are constrained by modality and genre at different levels of discourse, from the local level of individual clauses via interclausal packages and on to the topdown level of the text as a whole. Here, we examine how modality and genre interact in the development of the noun phrase, as a clause-internal linguistic constituent which plays an important role in the syntactic and informational complexity of a given piece of discourse as well as in expressing its general discourse stance (Ludo: Chapter 8 below).

Clauses are made up of syntactic phrases, in the form of verbal predicates (Ludo: Chapter 4) and their associated argument and adjunct nominals. These are local-level categories whose internal structure and patterning in combination with other phrases interact directly with the information structure of any given text. Together with the coordinating and subordinating inter-clausal linkages that define how densely clauses are organized into packages (Ludo: Chapter 5), clause-internal configurations constitute an important source of syntactic diversity and complexity (Berman & Ravid, 1999).

With age and additional schooling, phrase-level elements develop in length, complexity, and internal constituency (Scott, 1988). This increased complexity in the

structure and content of NPs, VPs, and APs is expressed in choice of head type, in longer, less linear structures, and in diversification of phrase structures (Cahana-Amitay, 1999; Ravid & Avidor, 1998; Ravid & Nir, 2000). The present paper focuses on the category of noun phrase as a clause-internal construction which contributes to both syntactic complexity and the expression of discourse stance. A pilot study on the distribution of NPs in Hebrew in narrative and expository texts written by children, adolescents, and adults showed that NP complexity develops in interaction with clause site (subject versus nonsubject) and text genre, while distribution of different types of pronominal and lexical NPs and degree of NP complexity proved diagnostic of age, clause site, and genre (Ravid, 2000).

Below we examine the patterning of NPs in spoken and written narrative and expository texts produced by gradeschoolers compared with adults in four languages: Dutch, English, Hebrew, and Spanish. Although speakers of these languages have access to different repertoires of formal options for NP-internal constructions and linkages (Myhill, 1992), we assume that the task of learning how to expand NPs in accordance with different discourse contexts will not differ in principle across these languages.

2. BACKGROUND RESEARCH ON NPS

NPs have been the subject of considerable attention in crosslinguistic and psycholinguistic studies of syntactic form / discourse function relations. They are central to the representation and processing of semantic, syntactic, and pragmatic information in discourse, because of the crucial role they play in the domain of reference in discourse (Ariel, 1990; Clark & Wasow, 1998; Hickmann, 1995; Lambrecht, 1994). In semantic terms, NPs are classified into participatory (or core) roles, such as agent and patient, versus circumstantial (or oblique) roles, such as benefactive, comitative and temporal relations. These roles are universal, but their overt coding features such as word order, case, or agreement marking, vary across languages (Andrews, 1985; Givón, 1993).

The morpho-syntactic classification of NPs includes types of xx NP structures (e.g., null elements, pronouns, lexical expressions), their grammatical relationships and clause sites (subject, object, nominal complement), type of modifiers (determiners, adjectives,

subordinate clauses), as well as their internal configuration or syntactic “architecture” and level of hierarchical complexity (Brendel-Viechnicki, 1999; Erdmann, 1988; Wasow, 1997). While these general dimensions are shared across languages, the particular types of NP selected to express various semantic roles in different grammatical sites as well as the internal structuring of NP constructions clearly depend on language-particular factors, such as the extent to which a language allows null arguments, the range of available pronominal expressions and the nature of noun modification (DuBois, 1987; Myhill, 1992).

As noted, much of the discourse-oriented research on NPs concerns their role in designating discourse entities in the general domain of reference (Van de Kopple, 1992). An important factor in this connection is the distinction between the two categories of information structure of identifiability and activation. *Identifiability* specifies whether a mental representation of a referent is already stored by the addressee, while *activation* refers to the speaker-writer’s assessment of a representation of an identifiable referent as already activated, and so “accessible” or not (Ariel, 1990; Lambrecht, 1994). Discourse referents with varying degrees of identifiability and activation states are encoded by different types of nominal elements, primarily either lexical or pronominal NPs. For example, NPs whose nuclei are proper nouns are easily identifiable, as are deictic pronouns. All but not only pronominal referents are active (that is, given), so that active referents may be realized by any type of NP – pronominal or lexical, definite or indefinite.

Different languages have different inventories of pronouns, and the discourse motivation for distinguishing between nouns and pronouns varies from one language to another, so that there may not necessarily be a universal or “objective” criteria for mapping activation status to grammatical NP types (Myhill, 1992:31-32). On the other hand, the major categorical distinction between lexical versus pronominal reference is critical in the kind of extended texts with which we are concerned, particularly considering the differences in thematic content and discourse stance we can expect between narrative and expository texts. Besides, although the pronominal resources of the four languages in this study -- Dutch, English, Hebrew, and Spanish – differ somewhat in scope and diversity, these

differences are not so extreme as to disallow comparability. The lexical / pronominal distinction thus occupies an important part of this study.

Another window on the interface of syntax and discourse is provided by the internal architecture of lexical NPs. These may consist of short, simplex structures consisting of a single lexical head, sometimes modified by one or more closed-class items (articles, demonstratives, possessives, quantifiers), xx and/or by an adjective. Examples of “simplex lexical NPs” includes Dutch *twee fietsers* ‘two cyclists’ [Jos, man, EW]², *de derde persoon* ‘the third person’ [Ron, man, SE], and the man’s name *Kyle* [Alfonso, boy, G, NW], or *these things* [Cyndie, girl, J, ES] in English. More complex lexical NPs include phrasal modifications, e.g., Hebrew *ha-keshet beyn anashim* ‘the-relationship between people’ [Tami, woman, EW] or Spanish *las envidias entre compañeros de clase* ‘the jealousies between companions of class = classmates’ [JUN, man, ES]. “Heavier” NPs consist of binomial constructions such as compounds, genitives, and appositionals, e.g., English *the issue of conflict* [James, boy H, EW] and *someone’s safety, someone’s well-being* [Barbara, woman, ES], Hebrew *matan hadraxa le-talmidim* ‘giving instruction to students’ [Sivan, woman, NS]. NPs may also have various kinds of clausal expansions, e.g., Spanish *actitudes que se han recogido en un centro* ‘attitudes that have been registered in a center’ [FJO, man, ES]. Longer and more complex NPs typically contain more than one lexical noun and contribute importantly to the overall information structure as well as to the syntactic complexity of the text as a whole.

Another important factor in the cognitive representation of discourse referents is the **syntactic organization** of the sentence, which provides clues to potential for activation. Specifically, the clause site of an NP in the sentence (as subject, object, or expressing some other syntactic relation) influences our perception of the activation state of its referents, mediated by the independent pragmatic factors of topic (what the sentence is about), focus

² In citing from the texts we elicited, we follow this convention: Subject’s (first) name, sex, age- or schooling group (G = gradeschool, J = junior high, H = high school, and man/woman for adults), text type (EW = Expository Written, ES = Expository Spoken, NW = Narrative Written, NS = Narrative Spoken). Ludo: would you please make an editorial decision here and apply across the chapters.

(novel component of a proposition), and relational categories between reference and proposition in given contexts (Du Bois, 1987; Lambrecht, 1994). Although topic and grammatical subject are distinct notions, lexical subjects in fact often represent the topic of the sentence or the “aboutness” of its information structure, by adding to the addressee’s knowledge about a previously established set of entities. Languages often express what the sentence is about by the grammatical subject relation (Keenan, 1976), serving metaphorically as “hitching posts for new knowledge” (Chafe, 1976). Our four target languages are comparable in this respect, since their grammatical subjects mark similar semantic and pragmatic roles in the types of texts that we have analyzed.

In **morpho-syntactic** terms, in many but by no means all languages, the relation of grammatical subject is directly identifiable by coding features such as word order, case marking, and agreement markers in the main clause (Andrews, 1985). The four languages in this study all mark subjects by certain of these coding features, most particularly by word order and inflectional marking of case and/or agreement. They all four use clause-initial word order to mark grammatical subjects, though to differing degrees. English has the strictest SVO order, e.g., xx *so the boy kept+up his incessant pestering* xx [[I changed the example from a personal pronoun subject, it is the 2nd English-speaking subject, also highschool, also a boy, also NW – you decide] *I ended up breaking his ruler* [Jay, boy, H, NW], while Dutch, Hebrew and Spanish although also basically SVO are more flexible in this respect. Hebrew uses SVO order in agent-oriented sentences, e.g., *yeladim mitraxakim exad me-ha-šeni* ‘kids drift-away from-one another’ [Shem, boy, J, ES], but predicate-oriented sentences have post-verbal subjects, e.g., *yesh draxim shonot liftor et ha-be’ayot* ‘be [=there are] different ways to-solve the-problems’ (Eylam, boy, J, EW), and Hebrew also allows VS order to depict presentative scenes and to topicalize information (Berman, 1990; Ravid, 1995). Spanish xx has even more flexible alternations of its basic SVO order, allowing VS order in many cases, and shifting the position of the subject NP in order to topicalize information, e.g., *una vez estaba yo en clase de tutoria* ‘once was I in tutoring class’ [SEC, boy, J, NW].

In general, then, subject NPs usually occur at the beginning of the clause in all four languages. As a result, the information contained in longer and more complex subject NPs is more difficult to process and to relate to other parts of the syntactic construction in which they occur: These more complex subject phrases may involve center-embedded constructions which impose an added cognitive burden on text producers and interpreters (Clark & Wasow, 1998; Ravid & Shlesinger, 1995).

Grammatical subjects in the four target languages also have varying degrees of inflectional agreement marking. English and Dutch have impoverished systems of inflectional morphology, e.g., English ‘*someone*, 3rd sing, *always walks*, 3rd sing, *away* [Christian, boy, G, EW] (Haeseryn, Romijn, Geerts, De Rooij & Van den Toorn, 1997; Quirk & Greenbaum, 1973). Hebrew and Spanish have richer inflectional systems of inflectional morphology; in both, nouns govern NP-internal agreement between head noun and its associated modifying elements (for number and gender in determiners and adjectives) and between subject nominals and their associated predicates (number, gender, and person in Hebrew, number and person in Spanish), e.g., Hebrew *sixa išit* ‘conversation,Fm personal,Fm’ Ludo: please specify how to enter grammatical information in glosses, maybe as in the Berman & Slobin volume [Yosi, man, NS]; Spanish *una buensma persona*, ‘a,Fm very good,Fm person,Fm’ [EVC, girl, J, NW].

Of the four target languages, Dutch and English generally require overt subject NPs, including expletives, while Hebrew and Spanish allow many subjectless constructions (Berman, 1980; Brucart, 1999; Haeseryn et al., 1997; Quirk and Greenbaum, 1973). Hebrew permits ellipsis of subject pronouns in simple clauses in past and future tense 1st and 2nd person, e.g., *ve-0-hayinu coxakim aleyha* ‘and-(we) would.1st Pl laugh.Pl at-her’ [Anat, girl, H, NW] and (like English) same subject ellipsis is optional in coordinate clauses and also (unlike English) in some subordinate clauses, e.g. *xx hu amar še-0 yavo* ‘he said that (he) would-come’. In Spanish, ellipsis of pronoun subjects is the default option in all simple clauses and in same subject coordinate and subordinate clauses,. e. g., *no creo que estas situaciones sean algo grave* ‘(I) not think these situations are something important’ [JLO,

boy, J, EW]. Spanish pronominal subjects are overtly marked only for contrastive emphasis, e.g., xx *pero la verdad es que hasta yo dudé* ‘but the truth is that even I had doubts’ [SEC, boy, J, NW] (Berman & Slobin, 1994: 540n).

3. SUBJECT NPS IN EXTENDED DISCOURSE

As noted, the focus of the current study is a syntactic analysis of NP types and structures in spoken and written text production in crosslinguistic and developmental perspective. Two aspects of NP type and structure are of specific interest to us. One concerns the kind of structure that occupies the subject NP slot under consideration – zero, pronominal or lexical NP. A second is the internal construction of subject NP categories, both lexical and pronominal. An innovative methodology for analyzing NP distribution in discourse was designed for Hebrew and then extended separately to Dutch, English, and Spanish, with language-specific adaptations defining NP types and structures for each language. Each of these four language-specific adaptations, which we term *NP Templates*, served as input for comparing subject NPs in our data-base, with the aim of ascertaining how choice of subject NP type and structure is a function of language, age, modality and genre.

In general, as revealed by converging evidence from all the chapters in this volume, the personal-experience narrative texts, relating to people and their interpersonal conflicts, are characterized by a subjective, specific, involved, immediate, affective, and personalized “discourse stance” (Ludo: see Chapter 8 below). In contrast, the expository texts, whose thematic content devolves around ideas and xx thoughts rather than events and activities, express a distanced, detached, cognitive, meta-textual, universalistic and generalized stance. These differences in stance are enhanced in the spoken medium, which promotes loose, even fragmented syntactic structure, and is governed by the restrictions of on-line processing. xx Writing, in contrast, is a medium which allows text producers to construct more complex structures, and to present, organize, and revise new information in texts without being similarly constrained by real time pressures of online processing (Ravid & Tolchinsky, xx in press; Strömqvist, 2000; Ludo: Chapter 2). We assume that these interrelated differences in stance find expression in choice of NP type and structure.

3.1 Type of Subject NP: Lexical versus Pronominal

Subject NPs in each of the four target languages come in three varieties: (1) *null or zero* pronominal subjects, e.g., Hebrew *0-ravti im xavera sheli* ‘quarreled:1stSg xx [=I] quarreled with a friend of mine’ [Reut, girl, G, NS], Spanish *0-estaba con otra chica* ‘was:1stSg with another girl’ [Inez, girl, J, NW]; (2) *pronominal* subject NPs, e.g., English *We each thought* [Charles, boy, G, NW], Dutch *Ik had veel pijn* ‘I had a lot of pain’ [Yaleesa, girl, G, NW]; and (3) *lexical* subject NPs, e.g., English *This pressure to do well may cause stress* [Chris, boy, H, EW], Hebrew *mekom avodati murkav ... me-anašim simpatiyim ve-yedidutiym* ‘my-place-of work is-composed ... of friendly and congenial people’ [Yoram, man, NW].

A major aim of this study is to determine to what extent the choice of subject NP type in texts produced by children and adults in each of the target languages is mediated by the modality and the genre of the context in which they are embedded.

We predict that in the four languages, spoken texts, constructed under the constraints of on-line production with a narrow processing window, will contain “lighter” subjects than written texts (Chafe, 1976, 1994). Also, narratives, xx as a genre that typically refers to sequences of events, will contain lighter subjects than expositives, which xx devolve around ideas (Berman, 2000a, in press; Britton, 1994; Halliday, 1988). Pronominal subjects are lighter than lexical subjects, since they refer to given, activated, and accessible discourse referents, and so require less effort and concentration in retrieval, since they do not represent new information (Chafe, 1994; Du Bois, 1987; Myhill, 1992). The idea of “lighter” subjects also includes shorter, simplex lexical NPs with no internal hierarchical structuring, since these do not contain much information in addition to the core noun, and they can be efficiently processed even within the narrow processing window of transient speech.

We thus expect fewer lexical subjects in spoken than in written texts, and in narratives than in expositives. We expect adult texts in general to be more lexical than children’s texts, reflecting the ability of mature speaker-writers to represent and process greater amounts of new information, to be better able to overcome the limitations of on-line

processing, and to take a less involved stance even in the course of personal narratives (Cahana-Amitay & Ravid, 1999; Ravid, 2000; Scholnick, Nelson & Miller, 1999). We do not foresee differences among the project languages in this respect.

3.2 Degree of Complexity

A second focus of this study concerns the joint effects of modality and genre on the internal structure of lexical subject NPs. All four languages have lexical NPs ranging from single core-NP forms through simplex structures modified by determiners and adjectives to more complex compound NP structures, to “heavy” NPs modified by PPs, NPs, complement clauses, relative clauses and infinitival complements. This paper focuses on one aspect of NP structural complexity, namely, *terminal and non-terminal* subject NPs. We defined terminal subject NPs as those containing a single lexical noun, e.g., English *one of the really bad conflicts* was with my really good friend Caitlin [Vanessa, girl, J, NW], Hebrew *yeladim tovim meshatfim et kol ha-yeladim* ‘children good:PL share with all the children = good kids let all the kids take part’ [Chen, girl G, ES]. Non-terminal subject NPs contain more than one lexical noun, e.g., Spanish *la visión del profesor* ‘the view of-the teacher’ [RAE, man, ES], English *addressing a conflict around a wedding is a touchy matter* [Sharon, woman, NW]. Non-terminal subjects thus typically contain phrasal or clausal modifications.

For all four languages, we predict more lexical subject NP complexity in written than in spoken and in expository than in narrative texts. We expect lexical subjects to become increasingly complex with age and schooling in both modalities and genres, aided by a larger lexical repertoire and more complete inventory of syntactic constructions combined with the mature ability to represent and process large syntactic segments (Anglin, 1993; Ravid, 2000; Scholnick et al., 1999).

3.3 Personal versus impersonal stance: pronominal subjects

A third aspect of NP syntax xx which we examined concerns the effect of genre and modality on the functional information conveyed by pronominal subjects in developing discourse. Pronouns may convey both personal and impersonal stance, and their classification depends on the pronominal resources available to a language in general. In the four languages in this

sample, personal and impersonal pronominal resources are determined by two factors: (a) the grammatical system and target language typology (the extent to which it requires a surface subject) and (b) discourse context – where pronoun reference may alternate between personal and generic in order to express an impersonal stance. All four languages distinguish deictic from anaphoric pronouns, and each language has specific pronominal forms for conveying generic, impersonal content. In Dutch this is, for example, the existential impersonal *er* as in *Er werd gescholden* ‘there was scolding’ [Janske, woman, EW]. English has expletive *it* and existential *there*, e.g. *It was a while back, in the summer* (Michael, boy, NS), *There was a big problem between my mom and dad* [Christina, girl J, ES]. Hebrew can but need not use an expletive pronoun *ze* ‘it, this, that’ as in *ze lo yafe še-0-osim davar kaze* ‘it (is) not nice that (people)-do,Pl such a thing’ [Shiran, f, G, EW]. Spanish relies more on subjectless constructions than on pronominal resources to express generic, impersonal stance. In addition to “empty” or “dummy” subject pronouns used to different extents in three of the four languages, all four languages can use personal pronouns with impersonal or generic reference for discourse purposes of a less personalized or deictic stance. This includes use of 2nd person pronouns (or inflectional person markers in Spanish and Hebrew) in all four of the languages, e.g. Spanish *en esta sociedad (tú) te dejas llevar* ‘in this society (you) let yourself go’ [ARA, woman, ES]; Dutch *je kan ruzie bijvoorbeeld oplossen* ‘you can for example solve an argument’ [Aimee, girl, G, EW]; and English *if you fight, you could get hurt* [Natalia, girl, G, ES]. Children thus need to learn not only the canonical deictic functions of pronouns in their language, but also how these same pronouns are used to express multifunctional options (Jisa & Viguié, 2000).

The prediction is that personal-narrative texts with their typically involved and subjective stance, will contain more personal than impersonal pronouns, while expository texts will contain relatively more impersonal than personal pronouns. We also expect spoken texts, produced with a present audience in mind, to express a more personal stance and so to include more personal pronouns than written texts, whose addressee is absent from the immediate context. These distinctions should be clearer in the texts of younger children than

in adult texts, in view of an increasing cognitive ability to take shifting perspectives on texts (Berman, 2000b). We also expect that these general trends will be affected by language-specific factors relative to the pronominal inventory in each language, as well as the richness of the inflectional system in each language (Ariel, 1990).

4. THE STUDY

In this initial examination of the distribution of subject NP types and structures in spoken and written narrative and expository texts, we focus on two age groups at the two extremes of our age range: gradeschoolers aged between 9 and 11 years, and university graduate adults, and on the four languages of Dutch, English, Hebrew, and Spanish. These four languages, like all those in our sample, are basically subject-initial and verb medial, though as noted earlier, they reflect differing levels of word order flexibility from English > Dutch > Hebrew > Spanish. This same ranking also applies to the extent to which they require surface subjects or not, interacting with the richness of inflectional marking for person in the four languages and whether they require or allow expletive or dummy subjects (again, yielding the order of: English > Dutch > Hebrew > Spanish). From these points of view, then, English and Spanish are typologically most distinct, with Dutch and Hebrew ranged between them.

4.1 Data-Base

The study population consisted of 20 participants per language (a subset of the total project population of 80 participants per language), in total 80 participants. In each of the four languages, 10 gradeschoolers and 10 adults were randomly selected from the original 20 subjects. The database for each language thus consists of 40 gradeschool texts and 40 adult texts, half spoken and half written, half narratives and half expository texts in each group, 320 texts altogether.

4.2 Analytical Procedures

Since texts are unequal in length, all our calculations are based on the internal distribution of the categories analyzed within NP totals in each text. Below we describe the step-by-step process we followed in analyzing distribution of subject NPs.

1. *Overall total: degree of lexicality.* In each text, the total number of subject NP slots was counted, classified by modality (spoken or written) and by genre (narrative or expository). The total number of what we termed *subject NP slots* thus consisted of all subject lexical nouns, all subject pronouns, and all zero subjects. The percentage of lexical subjects was then calculated out of this total to determine degree of text lexicality. Since only Hebrew and Spanish have subjectless constructions in simple clauses, we did not further analyze zero subjects in this study.

2. *Lexical NP analyses.* In each text, the total number of lexical subject NPs was counted and classified into *terminal* versus *non-terminal* NPs in accordance with the four language-specific NP templates. The following Hebrew example contains a non-terminal subject NP: *himan'ut mi-keshet karov im ha-sviva mafxita be'ayot ba'alil* 'avoidance of-(a) close connection with the-surroundings certainly reduces problems' [Benny, man, EW]. The percentage of non-terminal lexical subject NPs was then calculated out of the total number of lexical subject NPs.

3. *Pronominal subject NP analysis.* In each text, the total number of pronominal subject NP slots was counted and classified into personal and impersonal pronouns in accordance with the four language-specific NP templates. For example, the following English excerpt contains a first person plural subject pronoun used with a generically impersonal sense: *We have a lot to learn from everyone around us* [Jen, woman, EW]. The following English sentence contains two examples of personal subject pronouns: *She finally gave it to me after I asked lots of times* [Paul, boy, J, NW]. The percentage of personal and impersonal subject pronominal NPs was then calculated out of the total number of subject pronominal NPs to determine degree of (im)personal stance.

5. RESULTS

Results are presented in each section by age (gradeschoolers, adults), language (Dutch, English, Hebrew, Spanish), modality (spoken, written texts), and genre (narratives, expositories).

5.1 Subject NP type: lexical NPs

To determine the degree of text lexicality, we analysed the percentage of lexical subject NPs out of total subject slots (zero, pronominal, and lexical). A 2 (age: gradeschoolers, adults) x 4 (language: Dutch, English, Hebrew, Spanish) x 2 (genre: narratives, expositories) x 2 (modality: spoken, written) analysis on this measure yielded main effects of *age* ($F(1,72)=12.92$, $p<.001$; gradeschoolers $M=20.76\%$, adults $M=28.57\%$), *genre* ($F(1,72)=23.90$, $p<.0001$; narratives $M=20.25\%$, expositories $M=29.07\%$), and *modality* ($F(1,72)=39.51$, $p<.0001$; spoken $M=18.79\%$, written $M=30.53\%$).

The main effects were qualified by two-way interactions. Age interacted with language ($F(3,72)=4.61$, $p<.01$); the corresponding means are presented in Figure 1.

INSERT FIGURE 1 ABOUT HERE

Simple effects analyses confirmed that adults use more lexical NPs than 4th graders in Dutch, English and Spanish (all p 's $<.05$ or better), but not in Spanish. Age also interacted with genre ($F(1,72)=9.77$, $p<.003$); the corresponding means are presented in Figure 2.

INSERT FIGURE 2 ABOUT HERE

Simple effects analyses confirmed that adults use more lexical subject NPs in expository than in narrative texts ($p<.001$), while no such difference was found in the 4th graders. Finally, the interaction between genre and modality was also significant ($F(1,72)=8.30$, $p<.006$); the corresponding means are presented in Figure 3.

INSERT FIGURE 3 ABOUT HERE

Simple effects analyses revealed that expository texts contain more lexical subject NPs than narrative texts in the written ($p<.001$), but not in the spoken modality.

The remaining effects were not significant.

5.2 Non-terminal lexical subject NPs

To determine the degree of complexity of lexical subject NPs, we performed an analysis on the percentage of non-terminal subject NPs out of the total number of lexical subject NPs. A 2 (age: children, adults) x 4 (language: Dutch, English, Hebrew, Spanish) x 2 (genre: narratives, expositives) x 2 (modality: spoken, written) analysis on this measure yielded a main effect of *language* ($F(3,72)=8.08$, $p<.001$). A Tukey HSD post-hoc analysis showed that Spanish texts ($M=29.56\%$) contain more non-terminals than Dutch ($M=13.87\%$), English ($M=17.65\%$), and Hebrew ($M=15.52\%$) texts; all $p's<.01$. An *age* effect emerged as well ($F(1,72)=27.69$), $p<.0001$; gradeschoolers $M=12.57\%$, adults $M=25.73\%$).

The main effects were qualified by two-way interactions. Language interacted with age ($F(3,72)=2.73$, $p<.05$), see Figure 4.

INSERT FIGURE 4 ABOUT HERE

Simple effects analyses confirmed that the percentage of non-terminal NPs is higher in adults than in gradeschoolers in Hebrew ($p<.02$) and Spanish ($p's<.001$), but not in Dutch and English.

Genre and modality also interacted ($F(1,72)=4.34$, $p<.05$), see Figure 5.

INSERT FIGURE 5 ABOUT HERE

Simple effects analyses showed that written expositives contain more non-terminal NPs than do written narratives ($p<.02$), whereas spoken texts of both genres do not differ from each other.

The remaining effects were not significant.

5.3 Personal versus impersonal pronominal subjects

Finally, to determine the degree to which texts express personal versus impersonal stance, we examined the percentage of personal pronouns xx and of impersonal pronouns out of all subject pronouns.

A 2 (age: children, adults) x 4 (language: Dutch, English, Hebrew, Spanish) x 2 (genre: narratives, expositives) x 2 (modality: spoken, written) analysis on this measure yielded a main effect of *language* ($F(3,72)=23.59$, $p<.0001$). A Tukey HSD post-hoc

analysis showed that Spanish texts ($\underline{M}=27.35\%$) contain fewer personal subject pronouns than do Dutch ($\underline{M}=60.61\%$), English ($\underline{M}=55.14\%$), and Hebrew ($\underline{M}=55.11\%$) texts (all p 's $<.01$; no differences were observed among the three latter languages).

The analysis also yielded main effects of *genre* ($\underline{F}(1,72)=76.78$, $p<.0001$, narratives $\underline{M}=65.57\%$, expositories $\underline{M}=33.54\%$), and *modality* ($\underline{F}(1,72)=9.50$, $p<.003$, spoken $\underline{M}=54.07\%$, written $\underline{M}=45.03\%$).

Language interacted with modality ($\underline{F}(3,72)=3.56$, $p<.02$), as shown in Figure 6.

INSERT FIGURE 6 ABOUT HERE

Simple effects analyses confirmed that spoken texts contain more personal subject pronouns than written texts in English ($p<.02$) and Hebrew ($p<.003$), but not in Dutch and Spanish.

The remaining effects were not significant.

The proportion of Impersonal pronouns was calculated out of all pronominal subject NPs. A 2 (age: children, adults) x 3 (language: Dutch, English, Hebrew³) x 2 (genre: narratives, expositories) x 2 (modality: spoken, written) on this measure yielded a main effect of *language* ($\underline{F}(2,54)=18.26$, $p<.0001$). A Tukey HSD post-hoc analysis showed that English texts ($\underline{M}=9.45\%$) contain fewer impersonal subject pronouns than do Dutch ($\underline{M}=34.44\%$) and Hebrew ($\underline{M}=24.89\%$) texts (all p 's $<.01$; no differences were observed between Dutch and Hebrew). The analysis yielded two more main effects of *genre* ($\underline{F}(1,54)=28.19$, $p<.0001$; narratives $\underline{M}=12.45\%$, expositories $\underline{M}=33.41\%$), and *modality* ($\underline{F}(1,54)=17.30$, $p<.0002$, spoken $\underline{M}=28.03\%$, written $\underline{M}=17.82\%$).

Genre and modality interacted ($\underline{F}(1,54)=10.49$, $p<.003$), as shown in Figure 7.

INSERT FIGURE 7 ABOUT HERE

Simple effects analyses confirmed that both spoken and written expository texts contain more impersonal subject pronouns than narrative texts (p 's $<.004$); but while spoken expositories contain more impersonal pronouns than written ones ($p<.0001$), this difference does not apply to narratives.

The remaining effects were not significant.

6. Discussion

This is an exploratory study aimed at investigating the idea that, across languages, the type and structure of subject NPs is closely related to text genre and modality, and that its distribution changes between gradeschool and adulthood.

6.1 Summary of results

Our predictions are mostly confirmed by the results, as follows.

Lexicality of subjects. As predicted, the proportion of lexical subject NPs increases with age, although in only three out of the four target languages – Dutch, English, and Hebrew, and this increase occurs in expository rather than in narrative texts. The expository texts of adults, but not of gradeschool children, contain a higher percentage of lexical subjects than their narratives. Written texts contain a higher percentage of lexical subjects than spoken, and written expositives contain a higher percentage of lexical subjects than written narratives.

Complexity of lexical subjects. Our analysis here focused on the percentage of non-terminal subject NPs out of all lexical NPs. Languages differed somewhat in this respect, with Spanish showing a higher proportion of non-terminal subject NPs than Dutch, English, and Hebrew. Our age-related prediction in this respect was confirmed for Hebrew and Spanish, where the proportion of non-terminal lexical subjects increases with age. As predicted, written expository texts constitute the favored site for non-terminal subject NPs: they contain a higher proportion of such subjects than written narratives, on the one hand, and than spoken expositives, on the other.

Personal versus impersonal pronominal subjects. Our analysis focused on the percentage of personal compared with impersonal pronouns out of all subject pronouns. Contrary to our expectations, neither type of pronominal usage increases with age. And the languages differ in relative amounts of the two types of pronouns: Spanish texts contain a lower percentage of personal subject pronouns than the other three target languages, and

³ Spanish texts contained no free impersonal pronouns.

English texts contain a lower percentage of impersonal pronouns than Dutch and Hebrew. As predicted, narratives contain a higher percentage of personal subject pronouns than expositorys, and in English and Hebrew, spoken texts contain more personal subject pronouns than written. As predicted, expository texts contain more impersonal pronouns than narratives. Spoken expositorys contain more impersonal pronouns than do written expositorys.

6.2 Text lexicality and syntactic complexity

Text lexicality was defined by the occurrence of lexical subjects and of lexical subjects counted as non-terminal (i.e., with phrasal and clausal modification). More lexical subjects in a text means higher lexical density, that is, more content than function words; and also that more new information in the text is conveyed by those syntactic sites described in Chafe (1976) as “hitching posts for new knowledge”. Non-terminal subjects contribute even more to lexical density and to text informativeness, since each governs more than one lexical noun. Lexical, and particularly non-terminal, subject NPs also contribute to the degree of syntactic complexity and cohesion in the text, condensing semantic content in diverse hierarchical forms into a single syntactic phrase and making it highly informative, on the one hand, and more difficult to process, on the other. For example, Hebrew *hitnagšut reconot beyn pratim ba-xevra yoceret macav šel xikux ve-metax matmidim* ‘**(a) clash (of) desires among individuals in-society** creates (a) situation of constant friction and stress’ [Dganit, woman, EW] contains a non-terminal subject NP which condenses propositional information about desires clashing among individuals in society in one phrase, embedded as a subject in a clause which expresses another proposition, i.e., that such a clash leads to a certain social situation. Encoding this propositional information in a single NP in the syntactic position typically reserved for given (or active) information contributes to the degree of cohesion in the informational structure by linking previous and subsequent information.

Cross-linguistic differences notwithstanding, three factors affect subject NP lexicality and complexity in the texts which we examined: age, genre, and modality. Higher percentages of lexical, non-terminal subject NPs occur in adult than in gradeschoolers’ texts,

since adults are better able than children to represent and process information and to encode it in a densely packaged form. The preferred textual site for complex subject NPs is written expository texts. The expository genre is concerned with abstract concepts, processes, and ideas, encoded in lexical terms, while personal narratives refer to people pronominally. And the written modality allows for heavy subjects with expanded phrasal and clausal modifications, compared with the circumstances of spoken language, which make it difficult to assemble such large center-embedded phrases. Since subject position in SVO languages is particularly vulnerable to processing constraints as site for center-embedding, it is particularly difficult to realize this position with highly lexical and complex NPs. This makes the study of lexical subject NPs a rewarding domain for more general insights into text construction (Chafe, 1994; Clark & Wasow, 1998; Du Bois, 1987).

6.3 Personal / impersonal discourse stance

Unlike text lexicality, discourse stance (see 3 above and Chapter 8 below) as evaluated by amount of personal and impersonal pronouns seems to be in place from gradeschool and does not change as a function of age. From early on, the important factor that determines choice of pronoun type is genre: The objective, detached expository stance concerned with impersonal issues entails a higher percentage of impersonal pronouns such as Dutch *er*, English *it / there*, and Hebrew *ze*, while involved, subjective personal narratives employ more pronouns with a personal function. We found that a full array of language-specific pronouns is deployed by 4th graders with their appropriate functions, so that the pattern of their distribution does not change with age. In other words, gradeschoolers are as able as adults to convey discourse stance through the pronominal system in their languages.

4.4 Crosslinguistic comparisons

On the whole, our findings did not seem markedly affected by difference in genetic or areal origin of the four target languages (Semitic Hebrew versus Indo-European Germanic Dutch and English and Romance Spanish), nor by their typological character (highly

inflected Spanish and Hebrew versus weakly inflected Dutch and English). We did, however, find some language-specific patterns that can be attributed to differences in the pronominal repertoire and occurrence of subjectless constructions in the language. First, note that the proportion of lexical subjects in Spanish in this study is high to begin with and does not increase with age; Spanish texts contain more non-terminal lexical subjects than the three other target languages and fewer personal pronoun subjects. These findings are evidently because in Spanish “pro-drop” is the default in simple clauses and in same-subject coordinate and subordinate clauses, and it has no expletive pronouns or existential dummy subjects. As a result, Spanish exhibits far more subjectless constructions and far fewer uses of pronouns than the other languages – exactly in line with what was found for the oral narratives of children aged 3 to 9 years as well as adults in Spanish compared with English and Hebrew (Berman & Slobin, 1995, p. 540n). And there, too, the proportion of pronoun subjects in the other languages did not change significantly with age. The fact that English texts have fewer impersonal pronouns than Dutch and Hebrew may be because English pronouns are multifunctional, for example, *it* is both an anaphoric personal pronoun and an expletive, unlike Dutch *er* and Hebrew *ze*.

The crosslinguistic patternings revealed in the current study point to valuable directions for more in-depth study of different NP types, across larger populations, possibly with additional languages. Nonetheless, even this exploratory study of certain restricted facets of NP usage has confirmed certain major trends across our entire project: Genre distinctiveness is evident from the youngest age group in our sample, as shown by the fact that the distribution of personal and impersonal subject pronouns in narratives and expositives is stable across age groups. And the two extremes of text types in our study are represented by oral narratives at the one end, containing fewer lexical subjects and more personal pronouns, compared with written expository texts at the other end, with more, and more complex, lexical subjects, and more impersonal pronouns. In general, “heavy” lexical subjects are a hallmark of more detached, impersonal, objective, and informative written texts.

REFERENCES

- Andrews, A. 1985. The major functions of the noun phrase. In T. Shopen (ed.) *Language typology and syntactic description, Vol. I Clause structure*. Cambridge: Cambridge University Press, 62-154.
- Anglin, J.M. 1993. *Vocabulary development: a morphological analysis. Monographs of the Society for Research in Child Development* 58, 10.
- Ariel, M. 1990. *Accessing noun-phrase antecedents*. Routledge: London.
- Berman, R.A. 1980. The case of an (S)VO language: subjectless constructions in modern Hebrew. *Language* 56, 759-776.
- Berman, R.A. 1990. On acquiring an (S)VO language: subjectless constructions in children's Hebrew. *Linguistics* 28, 1135-1166.
- Berman, R.A. 1999a. Discourse stance. Opening paper in the panel "Talking and writing about conflict situations at different ages and in different languages", International Pragmatics Conference on Pragmatics and Negotiation, June 13-16, 1999, Tel Aviv University & the Hebrew University
- Berman, R.A. 2000a. Developing literacy in different contexts and in different languages. Final report submitted to the Spencer Foundation, Chicago. Spencer report.
- Berman, R.A. 2000b. Thematic perspectives on how children talk about interpersonal conflict. In M. Aparici, N. Algerich, J. Perera, E. Rosado & L. Tolchinsky (eds.). *Working papers in "Developing literacy across genres, modalities, and languages", Vol. III*. Barcelona: Institut de Ciències de l'Educació, Universitat de Barcelona, 65-76.
- Berman, R.A. In press. Genre and modality in developing discourse abilities. In C. Moder & A. Martinovic-Ziv (eds.), *Discourse across languages and cultures*. [Typological studies in Language series]. Amsterdam: John Benjamins.
- Berman, R.A. & D. Ravid. The oral / literate continuum: developmental perspectives. Final report submitted to the Israel Science Foundation, Tel Aviv University, September 1999.
- Berman, R.A. & D.I. Slobin. 1994. *Relating events in narrative*. Hillsdale, NJ: Erlbaum.
- Brendel Viechnicki, G. 1999. The role of noun phrases in the evidential systems of biology and linguistics. Paper presented at the the 41st Annual Meeting of the Midwest Modern Languages Association, Minneapolis, November 4-6, 1999.
- Britton, B.K. 1994. Understanding expository text. In M.A. Gernsbacher (ed.) *Handbook of psycholinguistics*. New York: Academic Press, 641-674.
- Brucart, J. M. 1999. La estructura del sintagma nominal: las oraciones de relativo, en V. Demonte e I. Bosque (eds.) *Gramática descriptiva de la Lengua española*. Madrid: Espasa Calpe.

- Cahana-Amitay, D. 1999. Developing verb structures across text types and modality. In R. Aisenman (ed.) *Developing literacy in different languages, genres, and modalities*. Tel Aviv University: Literacy Project, 142-157.
- Cahana-Amitay, D. & D. Ravid. 1999. Verbal vs. nominal reference to conflict situations: Evidence from Hebrew. Paper presented at the International Pragmatics Conference on Pragmatics and Negotiation, June 13-16, 1999, Tel Aviv University
- Chafe, W. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. Li (ed.) *Subject and topic*. New York: Academic Press, 25-76
- Chafe, W. 1994. *Discourse, consciousness, and time*. Chicago: The University of Chicago Press.
- Clark, H.H. & T. Wasow. 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37, 201-242.
- DuBois, B.L. 1982. The construction of noun phrases in biomedical journal articles. In Jorgen Hoedt et al. (eds) *Proceedings of the 3rd European Symposium on LSP: Pragmatics and LSP*. Copenhagen: The Copenhagen School of Economics.
- DuBois, J.W. 1987. The discourse basis of ergativity. *Language* 55, 59-138.
- Erdmann, P. 1988. On the principle of 'weight' in English. In C. Duncan-Rose & T. Vennemann (eds.) *On language: rhetorica, phonologica, syntactica: A festschrift for Robert P. Stockwell*. London: Routledge, 325-339.
- Gayraud, F. 2000. Le développement de la différenciation oral/écrit vu à travers le lexique. Thèse de Doctorat, Université de Lyon.
- Givón, T. 1993. *English grammar: a function-based introduction*. Amsterdam: Benjamins, 247-299.
- Haiman, J. & S.A. Thompson (eds.) 1988. *Clause combining in grammar and discourse*. Amsterdam: Benjamins.
- Hickmann, M. 1995. Discourse organization and the development of reference to person, time, and space in P. Fletcher & B. MacWhinney (eds.) *The handbook of child language*. Oxford, UK: Blackwell, 194-218.
- Haeseryn, W., Romijn, K., Geerts, G., De Rooij, J., & Van den Toorn, M. C. 1997. *Algemene Nederlandse Spraakkunst [Standard Dutch Grammar]*. Groningen: Martinus Nijhof.
- Halliday, M.A.K. 1988. On the language of physical science. In M. Gahdessy (ed.), *Registers of Written English: Structural Factors and Linguistic Features*. Kent, England: Pinter Publishers, 162-178.
- Jisa, H. & A. Viguié. 2000. A developmental view on the use of distancing devices in expository texts. Paper presented at the EARLI Special Interest Group Conference on Writing, Verona, Italy, September 7–9, 2000.
- Karmiloff-Smith, A. 1981. The grammatical marking of thematic structure. In W. Deutsch (ed.) *The child's conception of language*. London: Academic Press, 121-147.

- Katzenberger, I. In press. Expository text organization. *Discourse Studies*.
- Keenan, E. 1976. Towards a universal definition of subject. In C.N. Li, ed. *Subject and topic*. New York: Academic Press, 457-489.
- Lambrech, K. 1986. Topic, focus, and the grammar of spoken French. Ph.D. Dissertation, University of California, Berkeley.
- Lambrech, K. 1994. *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Mosenthal, P.B. 1985. Defining the expository discourse continuum. *Poetics* 14, 387-414.
- Myhill, J. 1992. *Typological discourse analysis. Quantitative approaches to the study of linguistic function*. Oxford: Blackwell. Chapter 2: Analyzing NP types pp. 20-52.
- Ochs, E. 1990. Cultural universals in the acquisition of language. *Papers and Reports on Child Language Development*, 29, 1-19.
- Quirk, R. & Greenbaum, S. 1973. *A Concise Grammar of Contemporary English*. London: Harcourt Brace Jovanovich.
- Ravid, D. 1995. *Language Change in Child and Adult Hebrew: A Psycholinguistic Perspective*. New York: Oxford University Press.
- Ravid, D. 2000. NP complexity in the development of text writing. In M. Aparici, N. Algerich, J. Perera, E. Rosado & L. Tolchinsky (eds.). *Developing literacy across genres, modalities, and languages*, Working Papers, Volume III. International Literacy Project. University of Barcelona, 163-170.
- Ravid, D. & A. Avidor. 1998. Acquisition of derived nominals in Hebrew: developmental and linguistic principles. *Journal of Child Language* 25, 229-266.
- Ravid, D. & M. Nir. 2000. On the development of the category of Adjective in Hebrew. In M. Beers, B. van den Bogaerde, G. Bol, J. de Jong, & C. Rooijmans (eds.) *From sound to sentence: Studies on first language acquisition*. Groningen: Center for Language and Cognition, 113-124.
- Ravid, D. & Y. Shlesinger. 1995. Factors in the selection of compound-type in spoken and written Hebrew. *Language Sciences* 17, 147-179.
- Ravid, D. & L. Tolchinsky. 2000. Developing linguistic literacy: A comprehensive model. In M. Aparici, N. Algerich, J. Perera, E. Rosado & L. Tolchinsky (eds.). *Developing literacy across genres, modalities, and languages*, Working Papers, Volume III. International Literacy Project. University of Barcelona, 1-29.
- Scott, C.M. 1988. Spoken and written syntax. In M.A. Nippold (ed.) *Later language development: Ages nine through nineteen*. 1st edition. Austin, Texas: PRO-ED, 49-55.
- Scholnick, E.K., K. Nelson & P. Miller (eds.) 1999. *Conceptual development: Piaget's legacy*. Mahwah, NJ: Erlbaum.

- Strömqvist, S. 2000. A note on pauses in speech and writing. In M. Aparici, N. Algerich, J. Perera, E. Rosado & L. Tolchinsky (eds.). *Developing literacy across genres, modalities, and languages*”, *Working Papers, Volume III*. International Literacy Project. University of Barcelona, 211-224.
- Thompson, S. 1978. Modern English from a typological point of view: Some implications of the function of word order. *Linguistische Berichte*, 54, 19-36.
- Van de Kopple, W. 1992. Noun phrases and the style of scientific discourse. In Stephen P. White, Neil Nakadate, and Roger D. Cherry (eds.) *A rhetoric of doing: Essays on written discourse in honor of James Kinneavy*. Carbondale: Southern Illinois University Press, 328-348.
- Wasow, T. 1997. Remarks on grammatical weight. *Language Variation and Change* 9, 81-105.

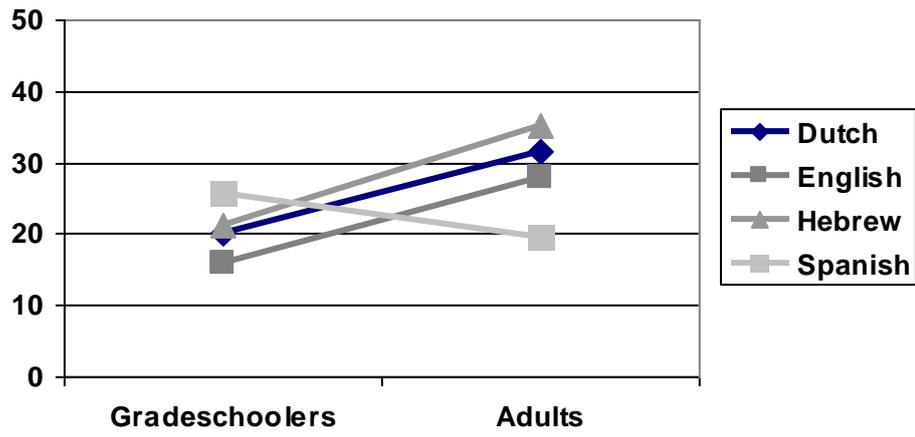


Figure 1. Interaction of age and language in percentage of lexical subject NPs.

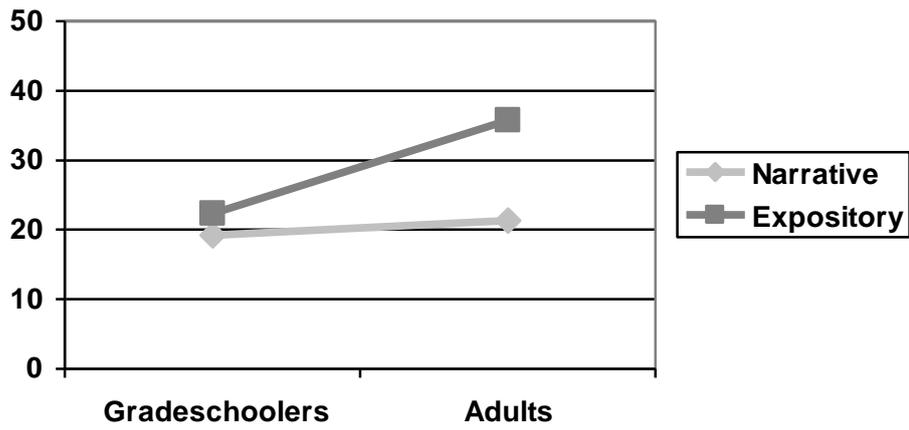


Figure 2. Interaction of age and genre in percentage of lexical subject NPs.

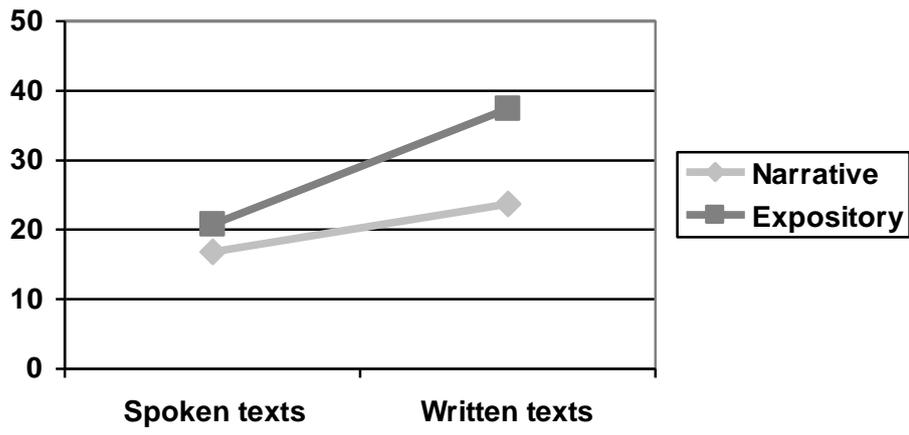


Figure 3. Interaction of genre and modality in percentage of lexical subject NPs.

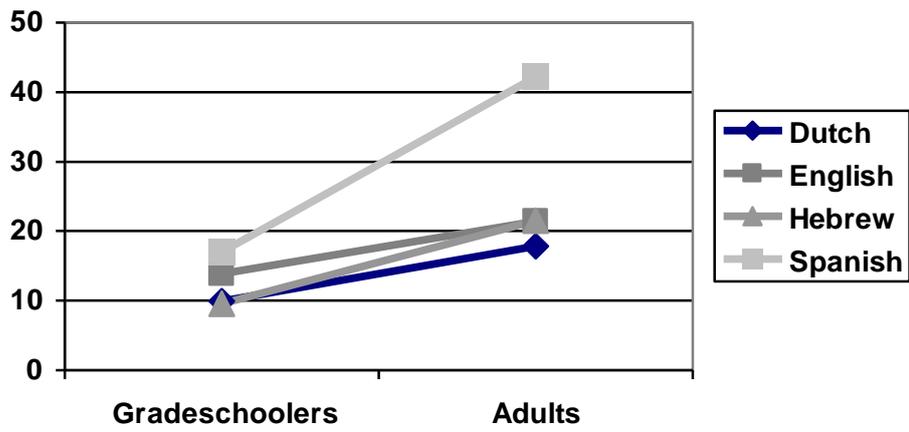


Figure 4. Interaction of age and language in percentage of non-terminal lexical subject NPs.

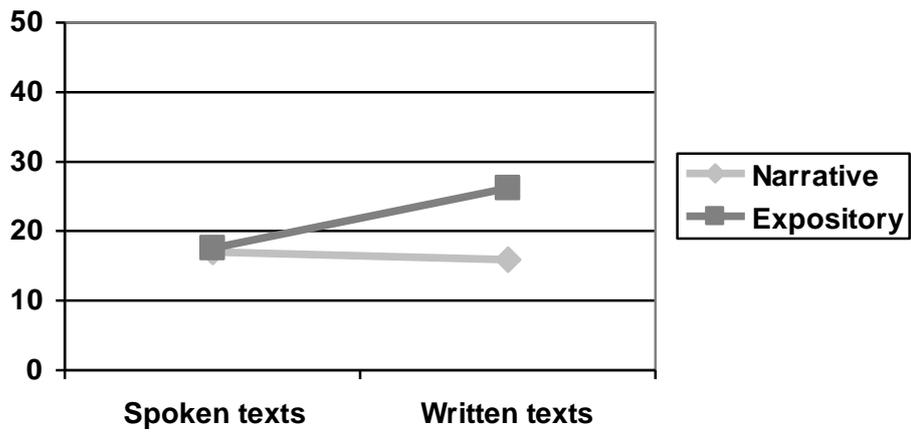


Figure 5. Interaction of genre and modality in percentage of non-terminal lexical subject NPs.

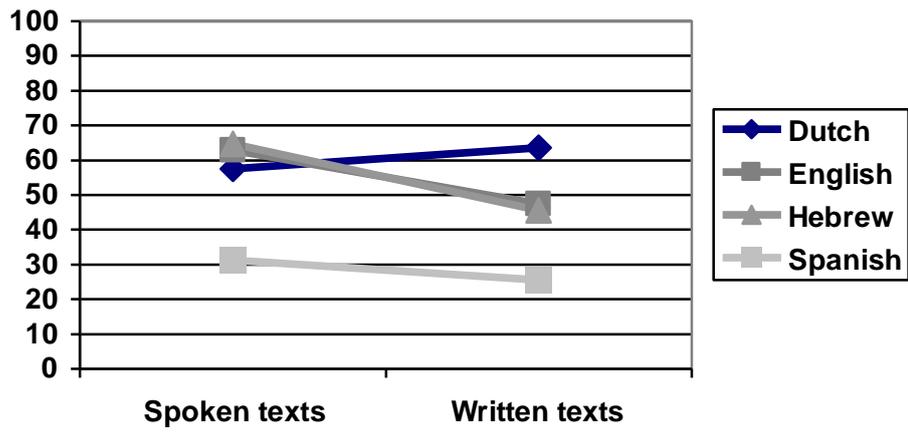


Figure 6. Interaction of language and modality in percentage of personal subject pronoun NPs out of all pronominal subject NPs.

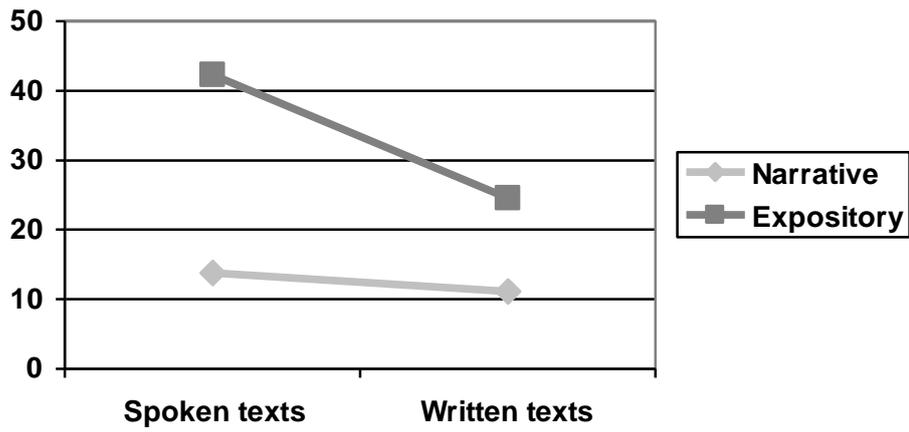


Figure 7. Interaction of genre and modality in percentage of impersonal pronominal subject NPs.

