August 4, 2000

Dorit Ravid and Shoshana Zilberbuch
Tel Aviv University
This is an initial version of the paper called "spoken and written text development"

Morpho-syntactic constructs in the development of spoken and written text production

Address for correspondence

Dr. Dorit Ravid
School of Education
Tel Aviv University
Tel Aviv 69978
Israel
doritr@ccsg.tau.ac.il

JCL, 2003

Abstract

1.0 Introduction: Moprho-syntax in later language development

Most researchers would agree that children growing up in a monolingual environment have access to the vast majority of morphological and syntactic structures of their language before they reach school age; nonetheless, a five-year-old hardly matches an adult or even a twelve-year-old in linguistic proficiency. Besides increasing vocabulary, one significant aspect of later language development is the ability to recruit different morpho-syntactic resources and to use them flexibly for diverse communicative purposes. The purpose of this paper is to examine knowledge and usage of

This usage can be investigated at a number of levels and from various points of view. One way to go about this investigation is to examine children's comprehension of later-developing morpho-syntactic constructs in isolated form or in text comprehension, or to get them to determine their grammaticality and / or appropriateness to well-defined contexts.

Complex sub-categorization in development

Analysis of morpho-syntactic constructions within their textual context may provide information about later linguistic development at two levels. Firstly, studying the usage of target constructions in authentic texts written by children and adolescents will demonstrate to what extent they are able to go beyond the merely "correct" yet decontexualized morpho-phonological production of these constructions to their appropriate usage in context. Secondly, this is a way to determine young writers' ability to distinguish text types and modalities through their usage of genre-appropriate morpho-syntactic forms. This is one goal of the current paper.

One domain which seems promising for the interface of morpho-syntax and text in development is the syntactic phrase. Clauses consist of syntactic phrases, specifically the predicate VP and satellite NPs. These are *bottom-up categories* whose internal structure, on the one hand, and the patterning of their combination with other phrases, on the other, produce differential clause-internal configurations. Phrases develop in length, complexity, and internal architecture with age and schooling (Scott, 1988). Studies which trace the development of VPs, APs and NPs indicate that increasing complexity is expressed in choice of head type, in longer and less linear structures, in various types of conjoining and subordination within the phrase, and in a diversification of phrase structures (Ravid & Avidor, 1998; Ravid & Nir, 2000). In this paper we focus on the internal structure of NPs, which permit an array of internal structures varying in complexity and configuration types. Initial studies have shown that written language, and especially expository texts, are characterized by complex lexical NPs which are crucial for the expression of hierarchical information structures participating in rhetorical text organization (Ravid, 2000; Tolchinsky, Perera, Algerich & Aparici, 1999).

Historically, bound compounds served as the Semitic highway to expressing a genitive or partitive relations, e.g., *beyt^ha-iš* `the man's house', *anfey^ha-ec* `the tree branches'. In Modern Hebrew, the genitive / partitive semantics is mainly expressed in analytical compounds, using the genitive particle *šel*, e.g., *ha-báyit šel ha-iš* `the house of the man', *ha-anafim šel ha-ec* 'the branches of the tree'. Bound N-N compounds with this semantics mainly occur in formal, higher-register written texts (Ravid & Shlesinger, 1995). But

N-N compounds constitute an interesting case of a historically well-entrenched Semitic construction, which does not occupy a privileged status as a lexical device in

early child Hebrew nor in any age group under experimental conditions, yet is used across the board in all facets of adult Hebrew.

Compounds mainly occur in early child-directed speech as lexicalized labels for familiar objects and events, e.g., *béged^yam* `bathing suit', *yom^hulédet* `birthday', and are often treated as single amalgams by young children, but they are not used for early lexical innovation in Hebrew.

Compounding may constitute a problem to Israeli children because of the morpho-phonological changes the bound stem undergoes, e.g., *bgadim / bigdey^kala* 'clothes / clothes^bride = bridal clothes'. Such changes diminish phonological faithfulness and render the relationship between free and bound stem opaque. However, stem changes characterize the whole nominal system and children are familiar with them from early on (Ravid, 1995a). By first grade, children perform most morpho-phonological stem changes correctly even in minor morphological groups (Levin et al., in press). Moreover, language change processes in Modern Hebrew have restricted stem changes in compounds and have rendered them less idiosyncratic and more regular, so that the distance between the way adults speak and what children have to learn is smaller (Ravid, 1995b; Ravid & Shlesinger, submitted).

Semantics may constitute the real problem for young learners, since sub-categorization requires manipulating the syntax-lexicon interface to create a complex hyponymic lexeme with one noun modifying another. In addition, Hebrew favors other, word-internal devices for expressing common relations such as agency, instrument and place, which occupy a privileged status in lexical word-formation (Berman, 1987). As for N-N compounds expressing the genitive relation, these are pre-empted by the analytic N-of-N construction (e.g., *ha-bgadim šel ha-iša* 'the clothes of the woman'), which is learned fairly early on.

Compounds expressing the genitive/partitive relation (e.g., *bigdey^ha-iša* ‘the woman’s clothes’), in contrast, do not emerge rather before schoolage, and occur primarily in poetry and literature as well as in high-register written texts. The sub-categorizing function of N-N compounds is clear to children by age four, and further development at gradeschool age is restricted to learning about their other functions, on the one hand, and to optimalizing morpho-phonological changes in the stem, on the other.

Denominal adjectives, a late historical development deriving from Biblical ethnic nouns which evolved in Medieval Hebrew into a full-fledged morphological class under the influence of Arabic

These changes are shared by all morphological operations on nominal stems in Hebrew, such as number and gender inflection, compound and genitive formation, as well as derivational suffixation[i] (Ravid, 1995a; Levin et al., in press).

Longitudinal observation yields a small number of examples where preschoolers produce unconventional N-A_{den} compounds. For example, Assaf (5;2) described a sports car (adult N-N compound *mexonit^sport*) by the N-A_{den} compound *óto spórti* `sportive car’. Levin et al., (in press) found clear development from kindergarten to first grade in producing N-A_{den} constructions (e.g., *halixa dubit* ‘walk bearlike = bearlike walk’), but children often supplied N-N constructions instead of the required N-$A_{den.}$ Moreover, N-A_{den} were found to be harder than inflected nouns with the same morpho-phonological stem changes.

Compare, for example, the specific meaning of *beyti* ‘domestic’ from *báyit* `house’ in *óxel beyti* ‘food domestic = homecooked food’ and in *tipus beyti* ‘type domestic = a stay-at-home type’. In some cases, the different shades in meaning are reflected in the two allomorphic variants of the denominal suffix: *i*- and *ani*. Compare, for example,

the two N-A_{den}s based on the noun *kol* 'voice' - *ta kol<u>i</u>* `lit. booth vocal = voice mail' and *dibur kol<u>ani</u>* 'speech loud = loud speech'.

There are cases where both are possible, with the compound expressing hyponymy and the N-A_{den} construction expressing 'N_1 with the property of N_2', e.g., compound *bad méši* `silk fabric' vs. N-$A_{den}$ *bad mešyi* `silken fabric'. However, not all N-N compounds and N-A_{den} constructions based on the same nominal behave so straightforwardly.

An examination of errors showed that children recognized the essential relationship between the two sub-categorization constructions, which often alternate in unpredictable ways in adult Hebrew: They occasionally interchanged them, producing N-N compounds instead of N-A_{den}s (e.g., *tiyul^šana* `year trip' for *tiyul šnati* `annual trip'), and N-A_{den}s for N-Ns (e.g., *xéder yalduti* `childish room' for *xadar^yeladim* `children's room'). This error type was especially prevalent in kindergarten, and children tended to exchange N-Ns for N-A_{den}s than vice versa (Ravid, 1997).

Once Hebrew-speaking children go beyond the typological tendencies of Hebrew towards forms with word-internal structures about age four, compounds are their first word-external device to express complex sub-categorization between two known lexical items.

The only developmental investigation of both constructions so far appears in Ravid (1997), who compared children's ability to produce N-N compounds and N-A_{den}s from kindergarten to first grade. This analysis showed that morpho-phonological knowledge of both types of constructions increased to the same extent during this period. However both age groups showed a better semantic grasp of compounding than of N-A_{den} formation. The results of this initial study indicate that the semantics of compounding is easier to acquire than that of N-A_{den}s. Indeed, N-Ns retain the

primary morphological function of complex sub-categorization in all types of adult discourse. In contrast, N-A_{den}s are first construed by children as serving a *syntactic* device of NP adjectival complementation, and are then learned as an alternative sub-categorization device. This happens as a result of linguistic and cognitive growth, which enables the child to perceive multiple functions of the same device, coupled with the massive exposure to written language and enhanced metalinguistic awareness as a result of the formal acquisition of literacy.

The Study

The current study focused on a comparison of complex nominal constructions in spoken and written text production in gradeschoolers, adolescents and adults.

2.1 Population

The study population consisted of 90 participants in 3 age groups: 1) 6th graders aged 11;5-13;1, mean age 11;11 (17 boys, 13 girls); 2) 11th graders aged 15;11-17;6, mean age 16;6 (13 boys, 17 girls); 3) Adults (students and college graduates) aged 22;8-49;2 (14 men, 16 women). All participants were monolingual Hebrew speakers from a middle-high socioeconomic background, living in the center of Israel.

2.2 Materials and procedure

Participants were asked to speak and write about two topics. One topic was biographical in nature, describing the life of a well-known public figure or of a friend or a family member. The second topic was expository in nature, and it was selected by the participant out of a given list of ten topics: *War, cinema, the city of Tel Aviv, clocks, cats, higher academic studies, football, the PLO, the zoo,* and *sculpture.*

Text elicitation was conducted in two stages. First, participants were interviewed orally on an individual basis by the second author. Each participant was

asked to select a biographical topic and tell about it, and then to select an expository topic and discuss it. The order of elicitation (biographical > expository) was intended to introduce and facilitate expository text production in view of the finding that development of text construction abilities emerges and consolidates earlier in narratives than in expository texts (Berman, 2000). These oral texts were recorded and transcribed on the same day of the elicitation.

In the second stage of text elicitation, participants were asked to write about the same biographical and expository topics they had discussed orally. The order of elicitation (spoken > written) was intended to facilitate text writing, since producing a written text requires planning, organization and revision processes, which may place a heavy cognitive burden on writers. Writing was done in the class forum. Participants were asked to write their compositions as though they were writing an encyclopedic entry, thus directing them towards using higher-register, literate, written discourse style. The 90 study participants each produced 4 texts, yielding altogether 360 texts: 180 biographical texts (half in speech, half in writing), and 180 expository texts (half in speech, half in writing). Appendix I lists the distribution of biographical and expository topics. These texts served as the basis for our analysis.

2.4 Analysis

In preparation for a statistical analysis, we examined two linguistic domains in the texts under investigation: 1) text size; 2) the complex nominal constructions under consideration – N-N compounds and denominal adjectives (A_{den}s).

2.4.1 Analysis of text size

Previous research has shown that text size increases with age. We measured text size in *words*, defined for Hebrew as graphemic sequences separated by spaces, and in *clauses*, as defined in Berman & Slobin (1994). Since texts produced by the

study participants were of different lengths, a third measure of *mean clause length* was introduced, defined as number of words divided by number of clauses. This measure has been shown to be a reliable diagnostic of textual development (Berman & Ravid, 1999; Berman & Verhoeven, to appear).

2.4.2 Analysis of nominal constructions

N-N compounds. N-N compounds were counted in three stages.

1. All tokens and all types were counted in each text. Compounds were lemmatized, that is, inflectional differences were disregarded.
2. Compounds were classified into those having *lexical* and *grammatical* heads. Grammatical compounds include a variety of function words as heads, e.g., *rov^ha-anašim* 'most^the people', *kol miney^dvarim* 'all sorts^things'. Grammatical compounds were eliminated from our analysis, since they do not testify to a complex and diverse lexicon.
3. All *lexical* N-N compounds were classified into two types: *lexicalized* and *novel* compounds. Lexicalized compounds are those rote-learned and retrieved by Hebrew speakers as idiomatic expressions or single lexemes, e.g., *beyt^séfer* 'Lit. house of books = school', as well as names of places, of books, of works of art, e.g., *drom^afrika* 'South Africa'. Six judges (all veteran teachers of Hebrew) judged compounds as either lexicalized or novel, following the procedure described in Ravid & Shlesinger (1995). The remaining syntactically-derived compounds were designated *novel*.

Denominal adjectives. Denominal adjectives were counted in three stages.

1. All tokens and all types were counted in each text. A_{den}s were lemmatized, that is, inflectional differences were disregarded.
2. All A_{den}s were classified into two syntactic types: *predicating* adjectives,

serving as predicate heads, e.g., *hu haya <u>šriri</u>* ‘He was <u>muscular</u>’; and *NP modifiers*, e.g., *séret ti’udi* ‘film documentary = documentary film’.

3. A_{den}s were also classified into those with *Hebrew* and *foreign* lexical bases, e.g., *yomi* ‘daily’ from *yom* ‘day’ (Hebrew-based A_{den}), *socialisti* ‘socialist’ (foreign-based A_{den}).

10% of all texts were analyzed by a separate judge (veteran Hebrew teacher) in addition to the two authors. 92% reliability was achieved, and cases of discrepancy were resolved by discussion.

2.5 Predictions

Given the earlier acquisition and the less restricted distribution of compounds versus A_{den}s, we predicted a higher number of compounds than A_{den}s in our texts.

For all measures, we predicted longer, more complex and novel structures in older age groups, in written rather than spoken texts, and in expository rather than biographic texts. This includes longer texts in terms of words and clauses, as well as longer clauses; an increased proportion of A_{den}s as compared with compounds (though compounds are always assumed to be more numerous); a higher proportion of novel vs. lexicalized compounds; a higher proportion of NP-modifying vs. predicating A_{den}s, as well as more foreign-based vs. Hebrew-based A_{den}s; and a higher proportion of NP-modifying A_{den}s vs. compounds as expressors of sub-categorization.

Development was predicted to be significant due to later language acquisition coupled with schooling resulting in more linguistically literate constructions. *Modality* was predicted to be significant due to the planned, organized and coherent nature of writing, which encourages revision, review and rewriting, allowing the retrieval of higher-register, literate lexical items and marked morpho-syntactic structures. *Genre* was predicted to be significant in both populations, given that expository texts are

lexically denser and syntactically more complex than narrative texts (Verhoeven & Berman, to appear).

3.0 Results

We first present the results for text size (3.1), for a comparative analysis of the distribution of N-N compounds vs. A_{den}s (3.2), and then for an analysis of N-N compounds (3.3), an analysis of A_{den}s (3.4), and finally a comparison of compounds and A_{den}s as expressors of complex sub-categorization (3.5).

3.1 Analysis of text size

We predicted a change in text size with age, modality, and genre. Table 1 presents the results for text size measured in number of words, number of clauses, and mean clause length.

INSERT TABLE 1 ABOUT HERE

To test our predictions we conducted a three-way ANOVA (age (3: 6th graders, 11th graders, adults) x modality (2: spoken, written) x genre (2: biographical, expository). On text size as measured in *number of words*, our predictions were confirmed. An effect for age emerged ($F(2,87)=10.61$, $p<.01$): Text size increases with age (6th graders M=138.52 < 11th graders M=176.49 < Adults M=246.22). An effect for modality emerged ($F(1,87)=166.69$, $p<.01$): Spoken texts are longer (M=265.14) than written texts (M=109.01). And an effect for genre emerged ($F(1,87)=5.19$, $p<.01$): Biographical texts are longer (M=193.85) than expository texts (M=180.3). These effects were mitigated by an interaction between age and modality ($F(2,87)=13.5$, $p<.01$). This interaction is presented in Figure 1.

INSERT FIGURE 1 ABOUT HERE

Figure 1 shows that in the two young age groups spoken texts have double the number of words than written texts, but adults' spoken texts have triple the number of words than their written texts.

On text size as measured in *number of clauses*, our predictions were partially confirmed. An effect for age emerged ($F(2,87)=13.61$, $p<.01$): Text size increases with age (6^{th} graders M=30.65, 11^{th} graders M=34.35 < Adults M=48.17). An effect for modality emerged ($F(1,87)=152.56$, $p<.01$): Spoken texts are longer (M=54.66) than written texts (M=20.79). However, no effect for genre emerged. These effects were mitigated by an interaction between age and modality ($F(2,87)=14.51$, $p<.01$). This interaction is presented in Figure 2.

INSERT FIGURE 2 ABOUT HERE

Figure 2 shows that in the two young age groups spoken texts have double the number of clauses than written texts, but adults' spoken texts have 3.5 the number of clauses than their written texts.

On text size as measured in *mean clause length*, our predictions were confirmed. An effect for age emerged ($F(2,87)=17.4$, $p<.01$): Clause length increases with age (6^{th} graders M=4.58 < 11^{th} graders M=4.9 < Adults M=5.02). An effect for modality emerged ($F(1,87)=39.69$, $p<.01$): Spoken texts contain *shorter* clauses (M=4.83) than written texts (M=5.32). And an effect for genre emerged ($F(1,87)=23.16$, $p<.01$): Biographical texts contain *shorter* clauses (M=4.89) than do expository texts (M=5.27). These effects were mitigated by an interaction between age and modality ($F(2,87)=9.07$, $p<.01$). This interaction is presented in Figure 3.

INSERT FIGURE 3 ABOUT HERE

Figure 3 shows that in 6th grade there is no difference in mean clause length between spoken and written texts, however in 11th grade and in adults clauses in written texts are longer than clauses in spoken texts.

A second interaction emerged between modality and genre ($F(1,87)=8.48$, $p<.01$). This interaction is presented in Figure 4.

INSERT FIGURE 4 ABOUT HERE

Figure 4 shows that there are no differences between spoken and written texts in mean clause length in biographical texts, however written expository texts contain longer clauses than do spoken texts.

3.2 A comparative analysis of the general distribution of N-N compounds vs. A_{den}s

We hypothesized that lexical N-N compounds are more basic and widespread than A_{den}s. In order to compare their amounts in texts of varying lengths, we compared the ratio of lexical N-N compound tokens versus A_{den} tokens out of the total number of clauses in each text. Clauses were selected as the baseline against which to compare the two study categories since compounds involve more than one word in Hebrew (Berman & Ravid, 1999). Table 2 presents the mean number of compounds and A_{den}s per clause by age group, modality and genre.

INSERT TABLE 2 ABOUT HERE

To test our hypotheses we conducted a four-way ANOVA of age (3: 6th graders, 11th graders, adults) x modality (2: spoken, written) x genre (2: biographical, expository) x nominal category (2: N-N compounds, A_{den}s) with repeated measures for modality, genre and nominal category. This analysis revealed an effect for nominal category ($F(1,87)=208.81$, $p<.01$): There were more N-N compounds per clause ($M=0.31$) than A_{den}s per clause ($M=0.10$).

We continued our analyses in each nominal category separately.

3.2.1 N-N compounds per clause

A three-way ANOVA (age (3: 6th graders, 11th graders, adults) x modality (2: spoken, written) x genre (2: biographical, expository) on the ratio of N-N compounds per clause showed that our predictions were partially confirmed. This analysis revealed an effect for age ($F(2,87)=11.35$, $p<.01$): The older the age group, the more N-N compounds per clause (6th graders M=0.22 < 11th graders M=0.31 < adults M=0.41). There was also an effect for modality $F(1,87)=149.95$, $p<.01$): There were fewer compounds per clause in spoken texts (M=0.19) than in written texts (M=0.43). An interaction of age and modality also emerged ($F(2,87)=12.90$, $p<.01$), depicted in Figure 5.

INSERT FIGURE 5 ABOUT HERE

Figure 5 shows that the older the age group, the larger the discrepancy in mean number of compounds per clause between written and spoken texts.

No effect for genre emerged in this analysis.

3.2.2 A_{den}s per clause

A three-way ANOVA (age (3: 6th graders, 11th graders, adults) x modality (2: spoken, written) x genre (2: biographical, expository) on the ratio of A_{den}s per clause showed that our predictions were partially confirmed. This analysis revealed an effect for age ($F(2,87)=10.80$, $p<.01$): The older the age group, the more A_{den}s per clause (6th graders M=0.25< 11th graders M=0.33< adults M=0.44. There was also an effect for modality $F(1,87)=158.57$, $p<.01$): There were fewer compounds per clause in spoken texts (M=0.21) than in written texts (M=0.47). An interaction of age and modality also emerged ($F(2,87)=12.56$, $p<.01$), depicted in Figure 6.

INSERT FIGURE 6 ABOUT HERE

Figure 6 shows that the older the age group, the more A_{den}s per clause in written texts, while in spoken texts this increase much milder.

No effect for genre emerged in this analysis.

3.3 Analysis of N-N compounds

To test our prediction about different patterning of novel compounds with age, modality and genre, we examined the ratio of *novel* (i.e., non-lexicalized) compound tokens out of all word tokens. Table 3 presents these measures by age group, modality, and genre.

INSERT TABLE 3 ABOUT HERE

A three-way ANOVA (age (3: 6th graders, 11th graders, adults) x modality (2: spoken, written) x genre (2: biographical, expository) on the ratio of novel N-N compounds per text (measured out of the total number of words) showed that our prediction was partially confirmed. No effect for age emerged. There was an effect for modality $F(1,84)=11.2$, $p<.01$): There were fewer novel compounds in spoken texts ($M=0.14$) than in written texts ($M=0.27$). No effect for genre emerged in this analysis.

3.4 Analysis of A_{den}s

To test our prediction about a different patterning of A_{den}s in the two syntactic sites with age, modality, and genre, we conducted two analyses.

3.4.1 Predicative versus NP-modifying A_{den}s

The first analysis focused on the internal distribution of predicative vs. NP-modifying A_{den} tokens, with the idea that this distribution would change in favor of NP-modifying A_{den}s with age, modality, and genre. However, this analysis could only be performed on 11th graders and adults, since there were only two 6th graders who produced A_{den}s in all four text types, while among the 11th graders there were 20

subjects and among the adults there were 25 who did so. Table 4 presents these measures by age group, modality, and genre.

INSERT TABLE 4 ABOUT HERE

A three-way ANOVA (age (3: 6th graders, 11th graders, adults) x modality (2: spoken, written) x genre (2: biographical, expository) on the ratio of A_{den} tokens out of all A_{den} tokens showed that our prediction was partially confirmed. No effect for age emerged. There was an effect for modality $F(1,43)=9.56$, $p<.01$): There were more predicative adjectives in the spoken texts (M=0.15) than in the written texts (0.06). No effect for genre emerged in this analysis.

3.4.2 Foreign- and Hebrew-based A_{den}s

The second analysis focused on the internal distribution of foreign- vs. Hebrew-based A_{den}s, with the idea that that this distribution would change in favor of foreign-based A_{den}s with age, modality, and genre. This analysis was again conducted only on 11th graders and adults who produced A_{den}s in all four text types. However, for this analysis we used types rather than tokens since we were interested in a lexical preference. Table 5 presents the ratio of Hebrew-based A_{den} types out of all A_{den} types.

INSERT TABLE 5 ABOUT HERE

A three-way ANOVA (age (3: 6th graders, 11th graders, adults) x modality (2: spoken, written) x genre (2: biographical, expository) on the ratio of A_{den} types out of all A_{den} types showed that our prediction was only partially confirmed. No effects for age, modality or genre emerged. Hebrew-based A_{den}s are definitely the majority across all age types. There was, however, an interaction of modality and genre ($F(1,43)=4.12$, $p<.05$), depicted in Figure 7.

INSERT FIGURE 7 ABOUT HERE

Figure 6 shows that written biographical texts contain a higher ratio of Hebrew-based A_{den}s (M=0.74) than all other text types.

3.5 Compounds and A_{den}s as expressors of complex sub-categorization

We have already presented results showing that (i) there are always more N-N compounds than A_{den}s in our study texts; and (ii) that there are more NP-modifying than predicative A_{den}s in all texts, and especially in written texts. We were further interested to see whether Hebrew users prefer N-N compounds to NP-modifying A_{den}s as primary expressors of complex sub-categorization in Hebrew. To find out the status of the two nominal categories, we compiled a new total consisting of all novel N-N compound types and all NP-modifying A_{den} types, and examined the distribution of the two nominal classes within this total, termed "complex NP". Note that for this analysis, too, we used types rather than tokens since we were interested in a lexical preference. We predicted a changing distribution in favor of all NP-modifying A_{den} types with age, modality and genre. Table 6 presents the distribution of the nominal classes in complex NPs.

INSERT TABLE 6 ABOUT HERE

A three-way ANOVA (age (3: 6th graders, 11th graders, adults) x modality (2: spoken, written) x genre (2: biographical, expository) on the ratio of N-N compound types out of all complex NP types shows that our prediction was confirmed. An effect for age emerged ($F(2,82)=12.54$, $p<.01$): Usage of N-N compounds decreases and that of NP-modifying A_{den}s increases with age (6th graders M=0.84 > 11th graders M=0.74, Adults M=0.74). An effect for modality emerged ($F(2,82)=4.51$, $p<.05$): Spoken texts contain a lower ratio of N-N compounds (M=0.76) than written texts (M=0.79). And an effect for genre emerged ($F(1,82)=7.52$, $p<.01$): Biographical texts

contain a higher ratio of N-N compounds (M=0.80) than do expository texts (M=0.75).

These effects were mitigated by an interaction of age and modality ($F(2,82)=4.08$, $p<.05$), depicted in Figure 8.

INSERT FIGURE 8 ABOUT HERE

Figure 8 shows that there were no modality differences in using N-N compounds in the younger age groups, while adults use more N-N compounds in written rather than in spoken texts.

4.2 Qualitative evidence for developmental changes

It is not only the statistical analysis that serves as evidence for on-going learning of complex nominals in children and adolescents. 6th graders' texts contained a variety of non-conventional N-N compounds and N-A_{den}s, indicating that by the end of gradeschool, complex nominals still constitute some challenge to Hebrew writers.

There were a number of morpho-syntactic violations in 6th N-N compounds, e.g., **maxalat le-nikayon* 'disease for-cleanliness' for *maxalat^nikayon* 'cleanliness disease', a hybrid form with a bound head and a PP for a head; a violation of NP word order in **tkufat^kol^ha-milxama* 'period^all^war' for *kol^tkufat^ha-milxama* 'throughout the war'; and N-N *nituxey^plastic* for correct N-A_{den} *nituxim plastiyi* 'plastic operations'. Other N-N compounds violated lexical and pragmatic conventionality (Clark, 1993), e.g., *šmot^mišpaxa* 'family names' for *šmot^bney^ha-mišpaxa* 'names of family members', *yoc'ey^ha-šoa* 'holocaust veterans' for *nicoley^ha-šoa* 'holocaust survivors'. Among the very few A_{den}s in 6th grade texts, violations involved structural ill-form: **mikre tragédi* for *mikre trági* 'tragic case', *xevratiyim* 'social,Pl' for *xevrutiyim* 'sociable, Pl', and N-A_{den} **nagif maxalati* 'diseasy virus' for correct N-N compound *nagif^maxala* 'disease virus'. 11th graders'

texts contained far fewer non-conventional constructions. There were only two violations of N-N compounding, both involving an overregulaized deverbal noun as compound head, e.g., *šlixat^ma'atafot* for *mišlóax^ma'atafot* 'sending envelopes' (see Ravid & Avidor, 1998, for discussion). A_{den} violations were slightly more numerous, e.g., **xayim politikayim* 'politicians' lives' for *xayim politiyim* 'political life', and N-A_{den} **sratim eymatiyim* 'horrific movies' for correct N-N *sirtey^eyma* 'horror movies'.

These examples indicate that while gradeschool children and, to a certain extent, highschoolers as well, are familiar with both study constructions, they still make structural over-generalizations, their knowledge of conventional lexical usage is still limited, the lexical status of specific complex nominals is not stable as yet, and in some cases, complex nominal constructions are not used in the text with full pragmatic appropriateness. This may seem slightly odd given the fact that 5-year-olds produce grammatical N-N compounds in free conversation and under experimental conditions, and first graders are well on the way to producing denominal adjectives under experimental conditions (Berman, 1987; Levin et al., in press). But producing complex nominals as labels for concepts during text composition is different from working within either a conversational framework or an experimental elicitation. During the top-down construction of a text segment, novice writers' attention is occupied with its thematic content and global text structure, as well as with the burden of constructing complex and complete clause packages. This process takes precedence over the bottom-up retrieval or assembly of complex concepts with their appropriately structured labels and their integration into the information flow. The result of this trade-off in novice writers is, as we have seen, ill-formed or inappropriate complex nominals, a vulnerable domain which has not been completely mastered by 6^{th} grade.

References

Berman, R.A. 2000. Final report to the Spencer Foundation. *"Developing Literacy in Different Contexts and in Different Languages"*, September 2000. Tel Aviv University.

Berman, R.A. & D. Ravid. 1999. The oral / literate continuum. Final project report for the Israeli Academy of Sciences. Tel Aviv University.

Berman, R.A. & D.I. Slobin. 1994. *Relating events in narrative*. Hillsdale, NJ: Erlbaum.

Berman, R.A. & L. Verhoeven. To appear. Introduction: Statistical measures of text size: A cross-linguistic analysis. In L. Verhoeven & R.A. Berman (eds.) *Written Language and Literacy* special issue.

Ravid, D. & Y. Shlesinger. 1995. Factors in the selection of compound-type in spoken and written Hebrew. *Language Sciences* 17, 147-179.

Modality		Spoken Texts				Written Texts			
Text Type		Biographical		Expository		Biographical		Expository	
Age	Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
6th grade	Words	192.27	85.97	200.00	85.45	92.63	39.12	69.17	37.00
	Clauses	42.93	17.97	43.1	19.25	21.67	10.59	14.9	8.22
	MCL	4.46	0.48	4.68	0.55	4.38	0.76	4.81	0.96
11th grade	Words	222.6	90.92	231.47	105.34	146.97	74.41	104.93	50.76
	Clauses	46.03	16.78	46.4	20.78	27.27	11.56	17.7	7.97
	MCL	4.81	0.69	4.99	0.66	5.36	1.00	5.92	0.97
Adults	Words	383.23	245.39	361.27	165.45	125.4	79.52	114.97	72.44
	Clauses	79.00	52.92	70.47	41.91	23.83	14.62	19.37	12.01
	MCL	4.94	0.58	5.1	0.62	5.36	1.05	6.11	1.3

Table 1. Means and standard deviations of text size measured in number of words, number of clauses, and mean clause length (MCL) measured by number of words divided by number of clauses, by age (3 groups), modality (spoken and written texts), and genre (biographical and expository texts).

Modality		Spoken Texts				Written Texts			
Text Type		Biographical		Expository		Biographical		Expository	
Age	Measure	Mean	SD	Mean	SD	Mean	SD	Mean	SD
6th grade	Compounds / Clauses	0.14	0.09	0.15	0.10	0.27	0.23	0.30	0.30
	A_{den}s / Clauses	0.19	0.11	0.16	0.11	0.33	0.23	0.33	0.30
11th grade	Compounds / Clauses	0.22	0.19	0.18	0.13	0.40	0.31	0.42	0.29
	A_{den}s / Clauses	0.25	0.19	0.19	0.13	0.45	0.29	0.44	0.29
Adults	Compounds / Clauses	0.25	0.17	0.18	0.14	0.62	0.29	0.59	0.36
	A_{den}s / Clauses	0.29	0.18	0.20	0.15	0.67	0.29	0.60	0.36

Table 2. Means and standard deviations of N-N compound tokens versus A_{den} tokens in a clause, by age (3 groups), modality (spoken and written texts), and genre (biographical and expository texts). Check with Gabi numbers in yellow – something is wrong there. How come mean A_{den} tokens is 0.10????

Modality		Spoken Texts				Written Texts			
Text Type		Biographical		Expository		Biographical		Expositc	
Age	Measure	Mean	SD	Mean	SD	Mean	SD	Mean	
6th grade	Novel compounds / Words	0.04	0.10	0.09	0.15	0.28	1.10	0.21	
11th grade	Novel compounds / Words	0.15	0.13	0.20	0.19	0.23	0.21	0.25	
Adults	Novel compounds / Words	0.18	0.11	0.18	0.15	0.22	0.12	0.43	

Table 3. Means and standard deviations of novel N-N compound tokens out of all word tokens by age (3 groups), modality (spoken and written texts), and genre (biographical and expository texts).

Modality		Spoken Texts				Written Texts			
Text Type		Biographical		Expository		Biographical		Expositc	
Age	Measure	Mean	SD	Mean	SD	Mean	SD	Mean	
11th grade	Predicative A_{den} tokens / All A_{den} tokens	0.12	0.26	0.12	0.25	0.05	0.13	0.08	
Adults	Predicative A_{den} tokens / All A_{den} tokens	0.17	0.24	0.19	0.25	0.03	0.08	0.08	

Table 4. Means and standard deviations tokens of predicative A_{den} out of all A_{den} tokens (both predicative and NP-modifying) by age (3 groups), modality (spoken and written texts), and genre (biographical and expository texts).

Modality		Spoken Texts				Written Texts			
Text Type		Biographical		Expository		Biographical		Exposito	
Age	Measure	Mean	SD	Mean	SD	Mean	SD	Mean	
11th grade	Hebrew A_{den} tokens / All A_{den} tokens	0.65	0.40	0.73	0.33	0.76	0.26	0.67	
Adults	Hebrew A_{den} tokens / All A_{den} tokens	0.65	0.28	0.63	0.27	0.73	0.36	0.57	

Table 5. Means and standard deviations tokens of Hebrew A_{den} out of all A_{den} tokens (both Hebrew and foreign-based) by age (3 groups), modality (spoken and written texts), and genre (biographical and expository texts).

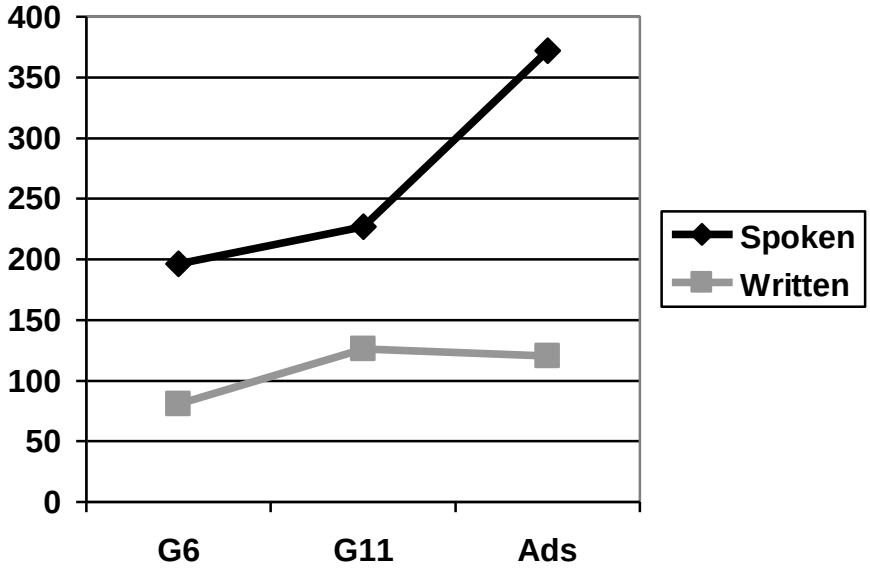

Figure 1. Interaction of age and modality in number of words

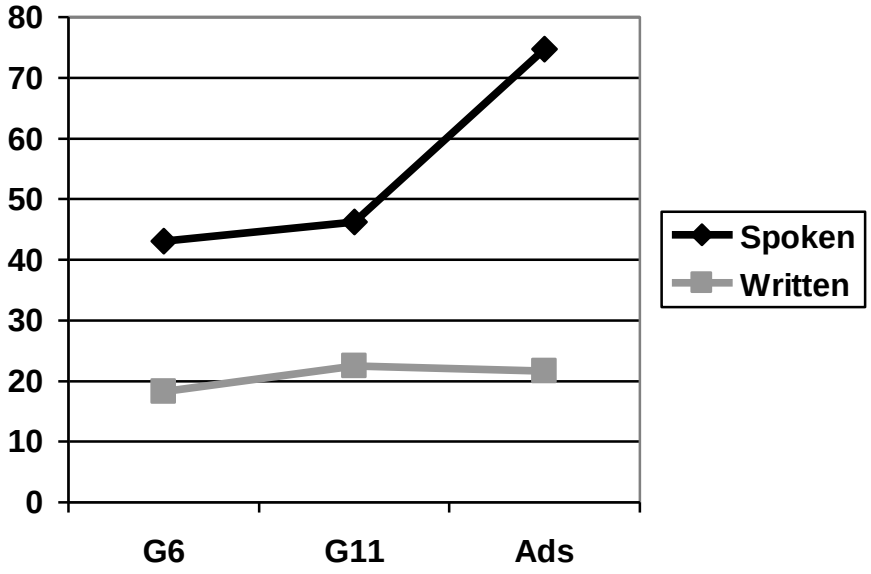

Figure 2. Interaction of age and modality in number of clauses.

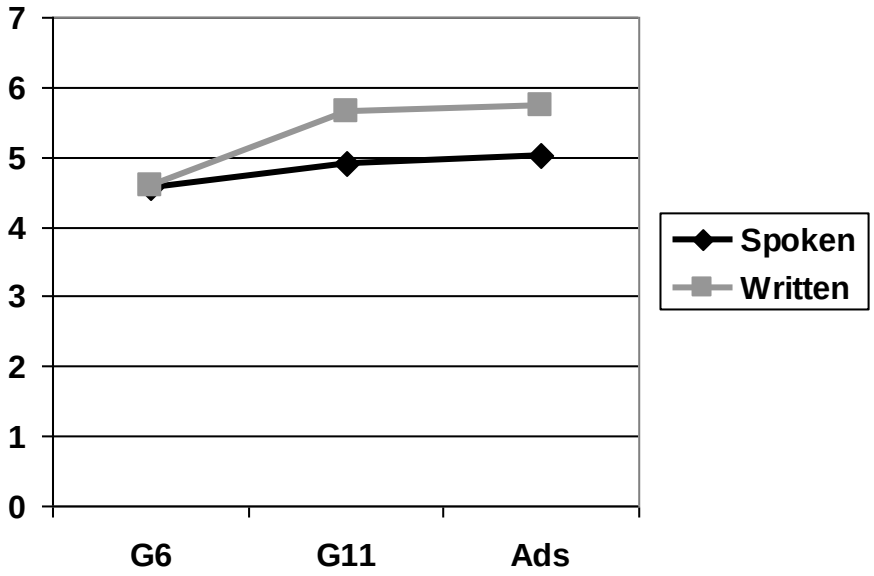

Figure 3. Interaction of age and modality in mean clause length.

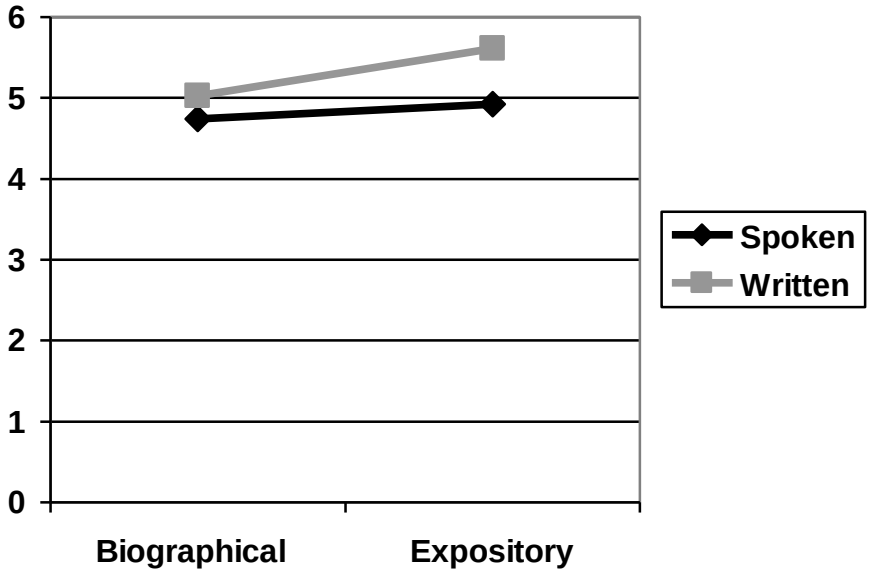

Figure 4. Interaction of modality and genre in mean clause length.

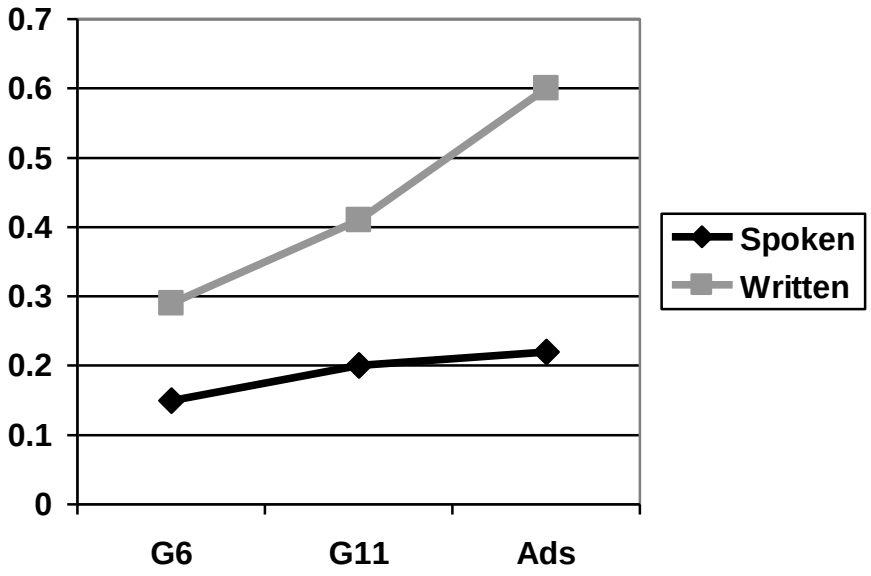

Figure 5. Interaction of age and modality in mean number of N-N compounds per clause.

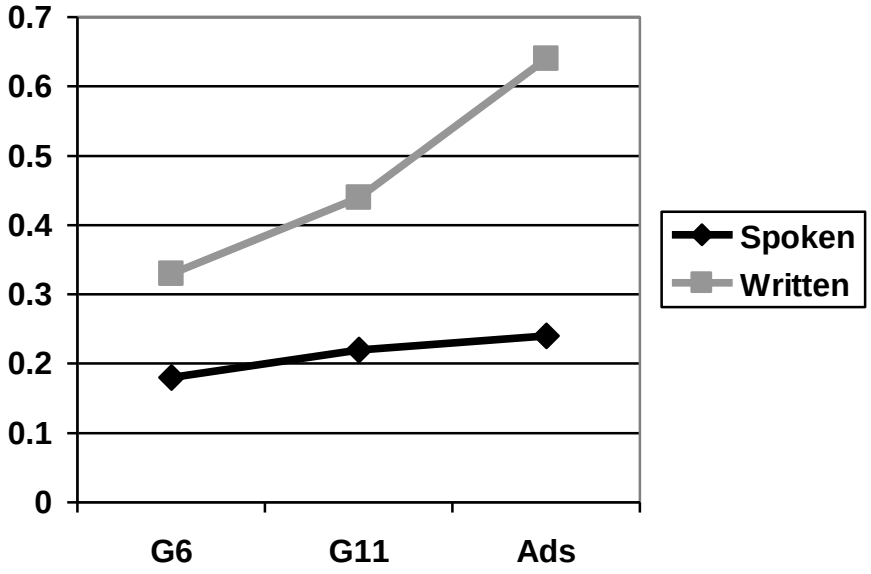

Figure 6. Interaction of age and modality in mean number of A_{den}s per clause.

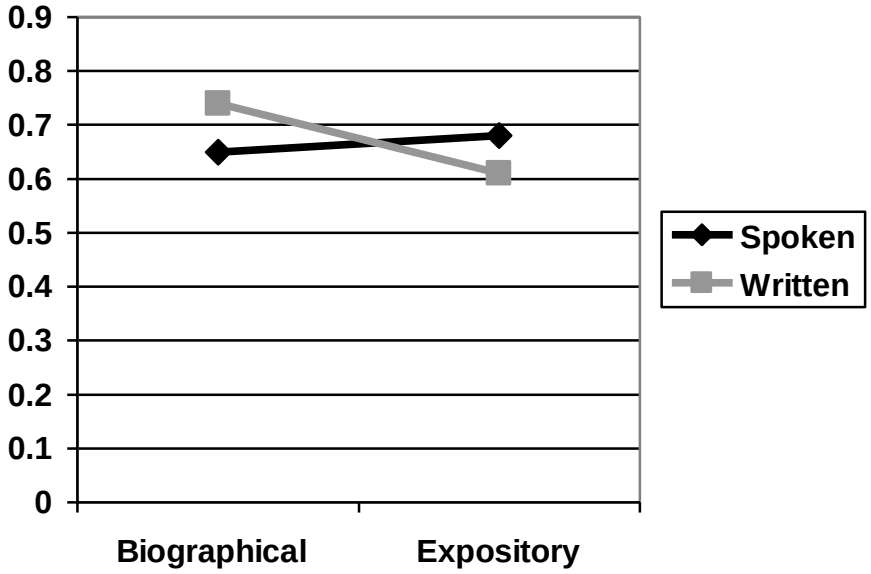

Figure 7. Interaction of modality and genre in number of Hebrew-based A_{den}s out of all A_{den}s.

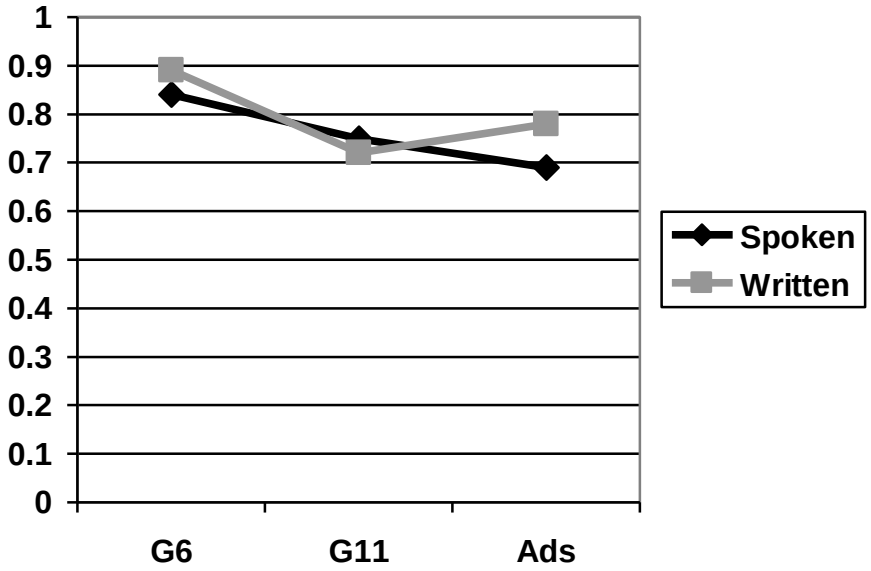

Figure 6. Interaction of age and modality in ratio of N-N compounds out of all complex NPs

Appendix I

1. Biographical topics

The life of…	***Proportion %***
Autobiography	24.4
Public figure	12.2
Father	17.8
Mother	8.9
Brother	7.8
Sister	6.7
Grandfather	2.2
Grandmother	4.4
Son	1.1
Daughter	1.1
Uncle	1.1
Aunt	1.1
Cousin	1.1
Friend (male)	3.3
Friend (female)	6.7

1. Expository topics

Topic	***Proportion %***
War	16.7
Cinema	16.7
Tel Aviv	2.2

Cats	7.8
Clocks	3.3
Universities	15.6
Football	12.2
PLO	6.7
Zoo	14.4
Sculpture	4.4

[i] In some cases suffix *–i* takes the allomorphic form of *–ani*, as in *rux-ani* (rather than *rux-i*) 'spirit-ual' from *rúax* 'spirit'.