Foundations of the Early Root Category:

Analyses of Linguistic Input to Hebrew-Speaking Children

Dorit Ravid[1], Orit Ashkenazi[1], Ronit Levie[1], Galit Ben Zadok[1], Tehila Grunwald[1,2],

Ron Bratslavsky[1,2], and Steven Gillis[3] *

[1]Tel Aviv University, [2]Israel Center for Educational Technology, [3]Antwerp University

Ravid, Dorit, Orit Ashkenazi, Ronit Levie, Galit Ben Zadok, Tehila Grunwald, Ron Bratslavsky and Steven Gillis. Foundations of the root category: analyses of linguistic input to Hebrew-speaking children. In R. Berman (ed.) *Acquisition and Development of Hebrew: From Infancy to Adolescence.* TILAR (Trends in Language Acquisition Research) series. Amsterdam: John Benjamins.

Supported by ISF Grant 285/13 to Dorit Ravid

Abstract

Researchers commonly consider the Semitic root to be the major lexical prime in Hebrew, relating morphological families in the major word classes. Psycholinguistic evidence supports the role of the consonantal root in acquisition and processing of Hebrew, from children's early ability to extract roots from familiar words to spelling and reading in Hebrew by adults. There is, however, little information regarding the actual distribution of roots in their canonical habitat of verbs in the Hebrew addressed to young children. To meet this lacuna, the authors analyzed verbs, roots, and *binyan* patterns in two types of linguistic input to children: (1) spoken -- child-directed speech to toddlers aged 1;8 -2;2 and (2) written – preschoolers' storybooks and 1st- 2nd grade texts. Findings include type and token frequencies of input verbs, distributions of full and defective root classes, morphological verb families, and semantic relations between verbs sharing the same root. The picture that emerges questions established views of root-based morphological families, and proposes a novel model of early verb and root learning in Hebrew.

Introduction

The mental lexicon of a literate adult is a richly interconnected and powerful system that provides the underpinning for semantic, conceptual, and linguistic knowledge (Mason & Just 2007; Vitevitch & Luce 1999). Words coexist in elaborate networks based on semantic-pragmatic, syntactic, morphological, and phonological properties, giving rise to systematic connections of varying strengths between items, which drive the emergence of categories and the ability to generalize beyond them in the creation of language-appropriate new words (Baayen 2010; Frishkoff, Collins-Thompson, Perfetti & Callan 2008). The encyclopedic nature of the mental lexicon is grounded in robust psycholinguistic evidence, including spontaneous malapropisms and tip-of-the-tongue and slip-of-the-tongue phenomena (Borodkin & Faust 2012; Fromkin 1980; Jaeger 2004), as well as experimental naming, lexical decision, and priming studies of word structure and meaning (Acha & Perea 2008; Clark 1996; Graves et al 2010; Katz et al 2012).

A major device organizing the mental lexicon of many languages is (derivational) morphology, which relates structural and semantic constituents within words (Haspelmath & Sims 2010; Marslen-Wilson 2007; Paterson, Alcocka & Liversedge 2011). In morphology-rich languages such as Hebrew, where many grammatical and lexical notions are encoded in word-internal structures (Ravid 2012), the morphological organization of the lexicon is paramount. The current chapter presents new data and analyses of the core morphological construct in Hebrew, the Semitic root, as a window on the early path of verb acquisition.

Root and pattern lexical organization

Most Hebrew scholars recognize the Semitic root as the major lexical device of Hebrew due to the pervasiveness of roots and root-based morphological families in the Hebrew lexicon (Berman

2012; Bolozky 1999; Nir 1993; Ornan 1990; Schwarzwald 2000). While linear formation constitutes an important part of Hebrew morphology (Ravid 2006), its lexicon is fed mainly by the non-linear combination of two sub-lexical morphological primes -- root and pattern (e.g., the root GDL 'grow' and the *hif'il* pattern together form the verb *higdil* 'enlarge'). The bulk of Hebrew content words, moreover, can be grouped into word families sharing a single root in different patterns (Berman 1987; Schwarzwald 2002). A *morphological family* such as that sharing the root GDL (e.g., in verbs like *gadal* 'grow', *higdil* 'enlarge', *gidel* 'raise'; adjectives like *gadol* 'big', *megudal* 'grown'; and nouns like *migdal* 'tower', *gidul* 'growth', and *gódel* 'size') shares a root both structurally -- the consonantal skeleton GDL -- and semantically -- the sense of 'grow'. While the basic form and reference of a morphological family are expressed by the root, patterns serve to form the word itself by providing the prosodic and phonological template into which roots are inserted, signaling categorial distinctions such as causative verb or action nominal. These morphological families -- sets of words in which patterns modulate the core meaning and structure of a given root across different syntactic classes -– permeate the Hebrew lexicon. All verbs are composed of root and verb pattern (*binyan*), and the nominal lexicon is also largely constituted of root-and-pattern structures that include some fifty different nominal patterns termed *mishkalim*, literally 'weights' (Avineri 1976; Ravid 1990; Schwarzwald & Cohen-Gross 2000). Much of the Hebrew lexicon is thus organized by *roots* connecting clusters of words with the same consonantal structure and basic lexical reference by *morphological patterns* grouping together words with the same prosodic structure and shared categorial class, for example, *CaCuC* passive resultative adjectives as in *šavur* 'broken', *katuv* 'written', *gamur* 'finished = done' (Berman 1994). This lexical organization by root and pattern morphology is evident in all Hebrew lexicons and dictionaries, which are arranged according to either root or pattern, in addition to orthographic principles (Avneyon 2007; Bolozky 2008; Even Shoshan 2003).

As a direct outcome of this ubiquitous non-linear lexical organization, Hebrew words readily lend themselves to the extraction of root skeletons, which, in turn, serve in new words. Root extraction is an extremely accommodating process which is fed by any and all word categories – content or grammatical, Hebrew or foreign, with or without internal morphological structure. For example, the abstract noun *hatrama* ‘priming’ is based on ŦRM, extracted from adverb *térem* ‘before’, the verb *kinsel* is based on KNSL, extracted from English *cancel*; while the adjective *me’udkan* ‘updated’ is based on IDKN extracted from the prepositional phrase *ad kan* ‘till here = so far’. The process identifies and extracts the consonantal skeleton of the base word, combining it with the appropriate category-assigning pattern, resulting in a word that conveys the meaning of the base word (or a facet thereof), much like zero derivation in English (Clark & Clark 1979).

The pervasiveness of Semitic roots in Hebrew lexical organization and new-word formation points to their central role in Hebrew lexical learning: Following repeated encounters with words, Hebrew-learning children pick up similarities in their internal components, leading to the identification and emergence of form-meaning categories representing the morphological relationships in the lexicon (Bybee 2006). Thus, Ravid (2003) reports abundant developmental evidence showing that gaining command of non-linear morphology is an essential component in learning Hebrew words (and see, too, Berman 1985, 2003; Clark 2003). Studies point to an early ability of Hebrew-speaking children to extract roots from familiar words and use them in novel forms (Berman 2000; Berman & Sagi 1981) -- for example, using the root TQN (from the verb *tiken* 'fix') and the agentive pattern *CaCaC* to coin novel *takan* 'fixer' for a technician or repairman. Current evidence points to the Semitic root as the most accessible Hebrew morpheme across different age groups and populations (Ben Zvi & Levie, this volume; Seroussi 2011) and to its role in adult word production, distinct from semantic and phonological factors (Deutsch & Meir 2011). The robustness of the root is evidenced even in contexts of language disability or environmental deprivation (Ravid, Levie & Avivi-Ben Zvi 2003; Ravid & Schiff 2006b; Levie, Ben Zvi & Ravid, submitted; Schiff & Ravid 2007). Written Hebrew enhances the status of the root as a core Hebrew morpheme: The diminished role of vowels, lack of marking of stop/spirant alternations, and the clear representation of the AHWY *matres lectionis* radicals all conspire to make written roots into continuous, clearly demarcated, visible entities, with the status of salient morphological units (Ravid 2012; Tolchinsky 2003). Studies of spelling show that Hebrew-speaking children rely on the distinction between root and function morphemes as early as 2nd grade (Gillis & Ravid 2006; Ravid 2001, 2005; Ravid & Bar-On 2005). Reading research in Hebrew and other Semitic languages demonstrates that words are linked through their roots rather than via their full forms (Bar-On & Ravid 2011; Ravid & Schiff 2006a; Schwarzwald 1981). According to Frost (2012), primes consisting of root letters consistently obtain reliable facilitation, no matter what the other letters in the word (Frost, Deutsch & Forster 2000; Velan, Frost, Deutsch & Plaut 2005).

Once the fundamental role of root and pattern structure in the Hebrew lexicon is recognized, a major question arises as to how it is learned. Acquisition of a large and variegated lexicon in childhood is critical for optimal language development and use in general, while for Hebrew-acquiring children, lexical learning involves accumulating and learning about consonantal roots, by and large in the context of a semi-productive derivational morphology that is fraught with lexical gaps. Consider the morphological family related by the root MSR 'deliver' convey', which contains the verbs *masar* 'deliver', *nimsar* 'be delivered', *hitmaser* 'devote oneself', the adjective *masur* 'devoted', and the nouns *mesira* 'delivery', *hitmasrut* 'dedication', *mesirut* 'devotion', *méser* 'message', *mimsar* 'relay', *tamsir* 'handout', *timsóret* 'transmission', and *misron* 'text message'. Young Hebrew speakers need to learn about the size and properties of each such family, the structural and semantic permutations of the root within this family, as well as the features of each of the words constructed from it. Understanding this learning job is important for insight into lexical and morphological development in Hebrew, and needs to be based on reliable knowhow as to the number and types of roots constituting the current Hebrew lexicon and how they relate to each other in morphological families -- information that is currently unavailable.

The present chapter represents an initial step towards obtaining such knowledge, by analyzing verb roots in the linguistic input to which young Hebrew-speaking children are exposed during the preschool years. The focus on input accords with usage-based approaches to language learning that view linguistic input as the major source available to children regarding the relative distributions and patterning of words and morphemes in their language (Hoff-Ginsberg 1985; Malsen et al 2004). Children have been shown to analyze the distributional properties of the ambient language, inducing linguistic structure from item frequency and from relationships between the properties of adjacent utterances (Behrens 2006; Goldberg, Casenhiser & White 2007; Theakston et al 2004). Input frequency has been associated with Age of Acquisition (AoA) effects, indicating that words acquired early on are processed faster in adult language than words acquired later (Brysbaert & Ghyselinck 2006; Juhasz 2005), as a factor that has implications for organization of the semantic system, word reading, and other lexical tasks in older language users, too (Johnston & Barry 2006; Turner, Valentine & Ellis 1998; Zevin & Seidenberg 2002).

Roots and patterns in verbs

Verbs were selected as the target category for analysis as the prototypical habitat of Hebrew root and pattern systematicity, in two senses. First, Hebrew verbs are assigned to one of seven conjugations termed *binyamin* (literally 'buildings') -- traditionally termed *pa'al* (*qal*)*, nif'al, hif'il, huf'al, pi'el, pu'al,* and *hitpa'el* – with each verb lemma a unique combination of a root in a specific *binyan* (Berman 1993a,b, 2012; Schwarzwald 1974, 1981).[1] A lesser-discussed facet of root and pattern structure is how it functions at the service of temporal inflectional paradigms -- typically marking Tense rather than Aspect (Berman 1978, in press; Berman & Dromi 1984). That is, unlike *mishkal* nominal patterns, what is traditionally known as '*binyan*' is not a single pattern, but actually an umbrella term for a bundle of temporal patterns which combine with the same root to construct a set of temporal stems unique in form to each *binyan,* expressing the temporal categories of Past, Present, Future Tense, Imperative Mood, and Infinitive form (Ashkenazi, Ravid & Gillis, submitted). For example, the verb meaning 'knit' from the root *s-r-g* is inserted into three different temporal patterns in *qal*, thus: Past Tense *CaCaC* > *sarag*, Present Tense *CoCeC* > *soreg*, and *PiCCoC* > Future *yisrog*, Imperative *tisrog*, Infinitive *lisrog* (where *P* stands for the temporal / person marking prefix);[2] while the verb meaning 'go/come-in, enter' is constructed from the root KNS inserted into two temporal patterns in *nif'al* as follows: Past and Present Tense *niCCaC* > *nixnas* and *PiCaCeC* > Future *yikanes, Imperative hikanes*, Infinitive *lehikanes*). Relating the different temporal patterns that make up a specific *binyan* is an essential part of constructing the temporal paradigm of the Hebrew verb, so that root construal is critical for learning about the basic inflectional structure of verbs.

A second, well established perspective on roots relates to their derivational role in the *binyan* system. Verbs cover the full range of semantic categories of activities, processes, and states, all formed in the small, closed system of *binyan* conjugations, with roots carrying the burden of lexical reference, to yield morphological verb families. For example, the root KNS 'go, come-in, enter' relates verbs in six different *binyanim*[3], as follows: *nif'al* > *nixnas* 'go, come-in, enter', *hif'il* > *hixnis* 'put in, insert', *huf'al* > *huxnas* 'be inserted', *pi'el* > *kines* 'assemble, Trans.', *pu'al* > *kunas* 'be assembled', and *hitpa'el* > *hitkanes* 'assemble, Intr.'; while the root ZMN 'invite' relates verbs in another set of five *binyan* patterns, thus: *hif'il* > *hizmin* 'invite, reserve', *huf'al* > *huzman* 'be invited, be reserved', *pi'el* > *zimen* 'summon', *pu'al* > *zuman* 'be summoned', *hitpa'el* > *hizdamen* 'happen, chance'. As these examples indicate, *binyan* conjugations are also associated with Verb Argument structures, relating them to syntactic-semantic valency functions such as causativity, inchoativity, reciprocity, reflexivity, and voice.

Consideration of different verbs in context makes it possible to form certain coarse-grained morpho-syntactic generalizations. First, *qal* is the syntactically and semantically most basic and pervasive *binyan* in all communicative contexts (Berman 1978, 1993; Berman, Naydits & Ravid 2011; Raz, in progress). Second, the *binyan* system is actually composed of two subsystems: (1) *qal*, *nif'al, hif'il* and *huf'al,* and (2) *pi'el, pu'al* and *hitpa'el*, each of which conveys the same set of syntactico-semantic valency functions (Ravid 2008). Thus, *nif'al* in the first set and *hitpa'el* in the second express low transitivity values (e.g., *nixnas* 'go, come-in, enter' and *hitkanes* 'assemble, Intr = get together'), while *hif'il* and *pi'el* convey highly transitive, often causative meanings and structures (e.g., *hixnis* 'put in, insert' and *kines* 'assemble-Trans, put together'). Apart from such broad generalizations, however, the *binyanim* constitute a semi-productive derivational system, so that the particular combination of specific verb patterns with specific roots is not predictable, nor can the specific shade of meaning of the verb be predicted from the said combination. In consequence, finer-grained information on verb lexical semantics needs to be learned for each combination in context.

It is thus safe to assume that root and pattern systematicity is mainly learned from verbs, with rich research evidence supporting this suggestion. We know that verbs constitute the first derivational system learned in early child Hebrew (Armon-Lotem & Berman 2003; Berman 1985, 1993,a,b). Recent analyses indicate that the earliest occurrences of root and pattern alternations in both child directed speech input and child speech output take place within the same *binyan*, most often based on the temporal stems of *qal* (Ashkenazi 2015). Verb structure, both temporal and derivational, is critical as the platform for later developments and acquisition of adjective and noun structure and meaning: About 40% of current Hebrew adjectives take the shape of Present Tense verbs in the form of the *benoni* present participle (Douani 2014), while deverbal nouns rely heavily on *binyan*-related morphology (Ravid & Avidor 1998).

Against this background, the present chapter aims to shed light on how early verbs can be acquired in Hebrew, as a language with a rich lexical texture in which verb systematicity plays an important role alongside of pervasive semantic and structural opacity. The focus in what follows is on verb roots, but our argument hinges crucially on the type of verb patterns into which roots are inserted. We will show that Hebrew-speaking children encounter mostly *semantically* coherent roots that transparently relate different a small number of root-sharing verb forms, while at the same time having to cope with prevalent structural opacity. This trend reverses with age: As they grow older, children encounter more, and more *structurally* transparent roots, while verb relationships become increasingly less semantically coherent as morphologically root-based families increase in size and complexity. This two-layered backbone of semantics and structure is what moves early verb learning forward.

The argument we present here is supported by new data and analyses of Hebrew verb input to preschool children, while relying on theoretical insights and empirical findings in the general psycholinguistic literature regarding initial semantic coherence and pervasive irregularity. Our expectations for small, semantically coherent verb families is based on studies (both computer simulations and empirical investigations) showing that complex morphological systems are learned by “starting small” (Elman 1993; Mariscal 2009), while they also make use of effective coherence markers[4] prominent in the system despite its apparent complexity when viewed from outside (Ackerman & Malouf 2013; Ravid 2011; Savickienė, Kempe & Brooks 2009). This theoretical approach has gained evidence in many studies in languages other than Hebrew showing that children work from specific to general in learning morpho-syntactic systems (Kuczaj & Maratsos 1983; Lieven et al 1997; Pine et al 1998; Rubino & Pine 1998; Rowland & Pine 2000); and that early constructions revolve around specific lexical items (Braine 1976; Tomasello 1992), which also tend to occur in only one aspectual form at early stages (Wilson 2003).

In the same way, structural irregularity and opacity are expected to prevail in input verb tokens, as in all languages morphological irregularities tend to reside in high frequency forms (Arnon & Clark 2011). For example, Davies (2009) showed that the 40 most frequent verb forms in American English are all irregular; while Rose, Stevenson & Whitehead (2002) found that just three irregular verbs (BE, HAVE and DO) accounted for fully a quarter of the attested verbs forms in the Reuters Corpus.

The chapter focuses on three related empirical questions as evidence to support these assumptions. First, what constitutes the verb and verb root lexicon in early Hebrew input? Second, what is the structural make-up of the roots in early input? And third, what is the nature of root-based morphological families in early input? The empirical part of the chapter is organized as follows. First, we present new data and analyses of verbs and roots across two sets of input corpora (Question 1); we then analyze the relevant roots in structural terms (Question 2); and examine morphological families and root semantics in the corpora (Question 3).

Data-Base

Our data-base consisted of both spoken and written samples, on the assumption that children from mid-high SES background like those in this study are exposed to linguistic input in both speech and writing. Research has shown that written language, as underlying literacy development, is lexically and syntactically denser and richer than spoken language, with more specific words covering more lexical domains (Berman & Ravid 2009; Biber 2009; Ravid & Tolchinsky 2002). Our chief interest here was in early child-directed input, where modality differences should be less apparent than in investigations of the linguistic usage of adolescents and adults. Nonetheless, we did expect parents to produce more fine-tuned speech with fewer different verbs and verb roots compared to expert-written child-directed materials, which also covered a relatively wide range of target ages.

In order to obtain as full a picture as possible of quantitative distributions and qualitative properties of verbs and root verbs in Hebrew input to young children, we made use of two novel corpora, spoken and written respectively. The *spoken* corpus consisted of the speech input of two sets of parents (mainly mothers) to their toddlers – one to a boy and the other to a girl (see Table 1). The two *spoken samples* were treated as a single corpus for present purposes, given the very high correlations that emerged not only between each child and his or her parents, but also between each parent and the other child, between the two sets of parents, and between the two children (Ashkenazi 2015). This yielded a corpus of 299,461 word tokens, including all language produced by the parents during the recording session -- spontaneous speech, repetitions, imitations, and onomatopoeia, songs, nursery rhymes, and stories – an inclusive policy based on the conviction that all linguistic input (and output) contributes to children's language learning.

The *written corpus* was merged from two samples of (1) children's storybooks targeted at preschoolers (that is, from toddlerhood to kindergarten, ages 1-6 years) composed or translated by writers of Israeli children's literature (Ben Zadok & Levie 2014) and (2) written school texts, primarily narratives, for beginner readers in 1st and 2nd grades (ages 6-7 years), written by child education experts (Grunwald 2014). We assumed here that written texts which target children enrich their language and provide input important to their language development (Cameron-Faulkner & Noble 2013). These two written corpora were merged for the current analysis to yield a corpus of 49,384 word tokens (see Table 1).

PLEASE INSERT TABLE 1 ABOUT HERE

Coding and Analyses

The coding of verb roots posed two challenges. First, we needed to demarcate the unit of '*verb*' in the context of temporal stems, which included the following forms: Past, Present, Future, Imperative, and Infinitive. Many Hebrew nouns and adjectives take the form of so-called *benoni* 'intermediate' participles in the form of zero derivation or syntactic conversion; for example, the form *šofet* could stand for either the *qal* verb '(he) judges' or the noun '(a) judge', *mefaked* stands for both the verb '(he) commands' and the noun 'commander', while *madrix* is either '(he) guides, leads' or the noun '(a) guide, leader' (Berman 1978, 1988). Since, however, all verbs in our corpora were analyzed as they appeared in context, spoken or written, we were able to exclude nominal forms from the analysis and to take into account only present-tense forms which were clearly either verbal , such as *lokáxat* '(is) taking, Fem', *mexabek* '(is) hugging', or adjectival, such as *ne'eman* '(is) loyal'.

Roots also posed a problem for analysis, since the same structural skeleton may have different meanings (e.g., SPR 'tell' and 'cut hair'). For present purposes, we defined 'root' as a unique structural skeleton relating verbs to one another, so that the number of roots in any given corpus corresponds to the number of different structural skeletons with similar morpho-phonological behavior. [5] This working definition leaves several semantic pitfalls in this respect, since, for example, a root like QRA may refer to reading, to calling out, and to calling by name. Yet this provided us with a necessary initial step, on the understanding that, once basic structural computations are in place, further semantic classifications can be implemented in the future on potentially polysemous items, both roots and verbs.

All verbs in the corpora were identified and coded for the following elements: Root, Root Type (full or defective, as defined below), *binyan* Pattern, Temporal stem (Past, Present, Future, Imperative, infinitive), and inflectional markers of Agreement (for Number, Gender, and Person). The analysis presented below disregards inflectional structure (Temporal stems and Agreement markers), in order to focus on verb lemmas and their (largely) derivational components – roots and *binyanim*. The question of Temporal stems is taken up again in the discussion at the end, in proposing a model of early verb and root learning in Hebrew.

Types and tokens

Since both type and token frequencies contribute to the emergence and entrenchment of linguistic categories, both were taken into consideration. For present purposes, *verb types* constitute verb lemmas, that is, a unique combination of root plus *binyan* yielding a verb. Thus, the combination of the root BWA with *qal* constitutes one verb type (citation form, 3rd person masculine singular past tense = *ba* 'come'), while the combination of the same root with *hif'il* constitutes another type (citation form = *hevi* 'bring').[6] Verb *tokens* were counted as all occurrences of fully inflected verb forms (e.g., *hevénu* 'brought, 1st, Plur - we brought'). *Root types* constitute different structural skeletons, so that BWA 'come' with a weak medial and final consonant is a different root type than, say, quadrilateral QLQL 'spoil' with reduplicated 1st and 2nd radicals. Root tokens consist of all the occurrences of the same root skeleton in the corpus. [7]

PLEASE INSERT TABLE 2 ABOUT HERE

Structural root categories

Table 2 presents the set of structural root categories defined for the current analysis. The first major category consists of *full* or *regular* (so-called "strong") triconsonantal roots in which all root radicals appear in every inflected or derived form, yielding transparent verb structures (e.g., *higdil* 'enlarge' in *hif'il,* based on the root GDL 'grow'). This category includes roots with pharyngeal and other 'gutturals' (so-called *groniyot* = 'made in the throat' in the Hebraic tradition) – here, referring primarily to radicals represented by the letters Ħ *chet,* I *ayin*, and H *heћ.* The second major category includes 10 main classes of *defective, irregular* ("weak") roots, primarily ones with non-consonantal radicals such as the glides, orthographic Y, W, or the glottal A (*alef*), sometimes also the weak radical N which may be deleted in some cases (Ravid, 1995; Schwarzwald 2013). These defective root categories effectively change the canonical verb structure (compare, for example, the *hif'il* form of the verbs *higdil* 'enlarge', *hirbic* from the full roots GDL, RBC respectively with the corresponding forms *horid* 'take down', *hipil* 'drop = cause to fall' from the defective roots YRD, and NPL respectively) and may result in word-final open syllables (e.g., *hevi* 'bring', *her'a* 'show' from the defective roots BWA and RAY respectively). Classification was initially based on the traditional division of Hebrew grammars by *gzarot* (structural root categories), with each instance of a root then re-examined for psycholinguistic evidence of its current morpho-phonological behavior within the verb system. For example, the root SBB 'turn around' was classified as defective (double) since, despite the transparent *qal* form *savav* 'walk in circles', forms constructed in other *binyanim* are non-canonical (e.g., in *pi'el*, *CoCeC* as in *sovev* 'turn,Trans' compared with regular *CiCeC* as in *sider* 'arrange' from the root SDR); on the other hand, the root ĦGG 'celebrate', traditionally defined as a double root, was classified here as a regular, full root, since it occurs only in transparent verb structures. Table 2 lists root categories in the traditional Hebraic order, from full to defective, from the initial to the final root radical position, with quadriliteral root categories containing more than the canonical three radicals further classified to reflect their internal structure.

Results are presented below for three analyses of verbs in early childhood input: (1) Distributions of verb tokens, verb types, and verb roots, (2) the breakdown of full and defective roots, and (3) morphological families of verbs deriving from a given root.

Results 1: Verb and root distributions in early childhood input

As noted earlier, a major lacuna motivating this study was the absence of empirical information on the root lexicon in Modern Hebrew. The analyses presented in this section constitute an initial step towards meeting this challenge. Table 3 provides information about verbs and verb roots in the two corpora.

PLEASE INSERT TABLE 3 ABOUT HERE

Verbs. Table 3 shows that the transcripts of spoken parental input contain five times as many verb *tokens* (about 55,000) as the written corpus of preschool children's storybooks and school texts (about 11,000). On the other hand, in line with our expectations for the contrast between interactive speech and monologic written texts, distributions of verb *types* reveal greater variety in the written materials: The spoken corpus contains about two-thirds the number of verb types as the written (684 vs. 1048 respectively). In other words, spoken parental input used 78 tokens per verb lemma, indicating substantial repetition of the same verbs; [8] in contrast, the written materials reflect a far lighter pattern of repetition, of 11 tokens per verb lemma, indicating that new verbs enter the corpus much more frequently. The two corpora combined present preschool children with altogether 1,186 different Hebrew verbs. Of these, 551 are shared by the two corpora, with another 133 occurring only in spoken parental input (e.g., *xafaf* 'wash hair' or *histarek* 'comb one's hair'), and another 497 only in the written materials (e.g., *ra'ad* 'tremble' and *hoda* 'thank').

PLEASE INSERT TABLE 4 (a,b) ABOUT HERE

Table 4 lists the forty most frequently occurring verbs in the two corpora, with their *binyan* patterns and respective frequency values. The two lists share 24 verbs (marked in blue), although in a different order of frequency, together displaying the 56 most frequent verbs in the study. The 10 most frequent verbs are shared by both lists, except for *sam* 'put', which occurs in the spoken. Interestingly, the twenty most frequent verbs constitute only 2.9% and 1.9% of verb types in the spoken and written corpora respectively, but they account for as high as 57% and 34% of verb tokens respectively – again underscoring the role of frequent verb repetition in spoken input.

Morphologically, the two lists are very similar: They consist mainly of verbs in *qal* (27 / 40 verbs in spoken input, 31 / 40 in written texts), with a few occurrences of *hif'il* (4 in both lists) and *pi'el* (7 and 4 respectively), and a single occurrence each of *hitpa'el* (the same verb meaning 'look' in both lists) and *nif'al* (those meaning 'take care" in the spoken input, and 'go / come in' in the written texts) in both lists.

The semantic-pragmatic breakdown of the lists is clearly linked to the genre and target audience of the two corpora. The five by far most frequent verbs in the spoken input are those expressing prototypical modality in shared actions - either lexically ('want'), or else in modal constructions inviting children to participate in joint activities, such as 'come', 'see', 'do ~ make' and 'bring' (Givón, 2009; Tomasello, 2003). The three most frequent verbs in the written corpus are those that mean 'say', 'see', and 'go' – expressing the three prototypical narrative functions of telling (description), understanding (interpretation), and locomotion (events). Both lists contain mental verbs (e.g., 'want', 'know', 'like ~ love', 'read', 'wait', 'be-careful', 'think', 'find'), *verba dicendi* ('say', 'sing', 'tell', 'talk', 'ask', 'tell', 'laugh', 'yell', 'answer'), verbs of perception ('see', 'hear', 'look', 'hurt'*)*, and (change of) state or unaccusative verbs ('fall', 'happen', 'sleep') – with written texts having more members in the clearly cognitive categories. Both lists contain activity verbs ('do', 'put', 'eat', 'take', 'give', 'show', 'help', 'draw', 'open', 'drink', 'close', 'bring', 'cry', 'play', 'look (for)', 'hold', 'fix', 'jump') and self-initiated motion verbs ('come', 'go, come-out', 'walk', 'go back', 'run', 'arrive', 'go, come-in') – with more motion verbs in the spoken input. Only the written list includes a common aspectual verb 'begin'.

Roots. Analysis of the breakdown of roots again demonstrates the highly repetitive nature of parental speech as against the more variegated lexicon of written texts. Thus, despite its much greater size, the spoken corpus contains only 70% of the number of different roots in the written corpus (521 vs. 744). Further, the ratio of verb types to roots shows the spoken transcripts to contain 1.34 verb types per root as against 1.41 verb types per root in the written texts. Since roots are the basis of shared morphological families, this means that written texts contain more and/or larger morphological families. Together, the two corpora contain 816 different roots, 455 of which are shared, for example, BNY ‘build’, YCA ‘leave’, SDR ‘fix, put in order’, and QLP ‘peel’. The spoken parental corpus has 69 other roots not shared by the written corpus (for example, HRG ‘kill’, QYA ‘vomit’), while the written corpus contains 292 additional roots not shared by the spoken corpus (e.g., AĦL ‘wish’, MCMC ‘blink’). While the combined figure of 816 clearly does not represent the total number of roots in Modern Hebrew, besides needing to be further analyzed for semantic ambiguities, it does define the order of magnitude we are looking at in input to children, and it gives a rough idea of the roots of lexical reference they experience. [9] When we compare the figures of 1,186 different verbs and 816 different roots, it seems that the majority of verbs children hear (and read in the early school years) do not share roots, with 1.45 roots per verb lemma. We elaborate on this question in the discussion section.

Results 2: Structural root categories in early childhood input

A second perspective on roots is structural in nature, inquiring into the distributions of full and defective roots in the two corpora. Figure 1 provides the information on root tokens and types in the corpus of spoken parental input.

PLEASE INSERT FIGURES 1 (a,b) AND 2 (a,b) ABOUT HERE

Despite the different sizes of the two corpora, the root-type distributions are extremely similar, pointing to the fundamental make-up of the Hebrew root lexicon. In both the spoken and written corpora, 64% of the roots are full or non-defective, that is, they consist of three consonantal radicals that show up in every verb form constructed out of them, yielding canonical, transparent temporal stems (e.g., BDQ 'check, examine'). Both corpora contain exactly 9% quadrilateral roots (mostly reduplicated, such as GLGL 'roll'), which are considered non-defective in the Hebraist tradition.[10] Thus, close to three-quarters of the root types in the linguistic input to children are full. Of the defective roots, over 15% of the irregular root types in both corpora come from two main classes -- *y*-final (e.g., GLY 'find out, discover') and *w/y* medial (e.g., RWC 'run') -- while four lesser classes – *y*-initial (YDIʼ 'know'), *n*-initial (NPL 'fall'), *?*-final (MCA 'find') and double (SBB 'turn around') contribute another 9% to both pies in similar proportions across the two corpora.

The input corpora differs from the written materials largely in *token* distributions, consisting of slightly over one-quarter full triconsonantal roots, with hardly any quadriliterals. The overwhelming majority of root tokens in the spoken input are defective, as predicted, with a huge group of *y*-final roots which, together with a small percentage of *?*-final roots, account for over one-quarter of the root tokens, largely due to the prevalence of RCY 'want', RAY 'see', and ISY 'do'. Roots with medial *w/y* make up another 20% of the spoken sample, largely due to the prevalence of BWA (a composite root) in the root-related verbs meaning 'come' and 'bring'. Defective categories such as *?*-initial (e.g., AKL 'eat'), *y*-initial, *n*-initial and their composites occupy a much larger slice of the token pie than their size in the type lexicon, constituting the final one quarter of the token distribution.

In contrast, over 45% of the root tokens in the written corpus are full (and quadrilateral), with fewer than one-quarter in all made up of *y*-finals, *w/y* medials and their composites. Here, too, *?*-initials, *y*-initials, *n*-initials and their composites are inflated beyond their type size, due to the prevalence of specific roots such as AMR 'say' and YŠB 'sit, reside' in this corpus.

To sum up the findings reviewed in this section, most parental token input consists of defective roots, as does 55% of written input. As dramatic illustration of this point, the twenty most frequent verbs in the spoken input all have defective roots, as do almost all of the twenty frequent verbs in the written corpus (Table 4). This distribution of the defective roots is to be expected, given what we know about how irregularities are maintained over time across languages (Arnon & Clark 2011; Davies (2009). In contrast, full, regular roots occupy most of the root *type* lexicon in both spoken and written input, meaning that most of the new roots children are exposed to are full. This becomes more apparent when considering the increasingly regular root structures in the second half of Table 4 for both spoken and written input. Figure 3 shows that as children grow older, parental input contains many more full roots, while the stock of early defective roots is exhausted in token occurrences and is not replenished. Thus, full roots carry the burden of lexical learning in Hebrew.

PLEASE INSERT FIGURE 3 ABOUT HERE

Results 3: Root-based morphological families in early childhood input

A final perspective on early verbs in child-directed linguistic input concerns the role of roots in relating verbs in morphological families across different *binyamin.* To begin with, consider Figures 4 and 5 for the distribution of the seven *binyan* verb patterns in the spoken and written input corpora.

PLEASE INSERT FIGURES 4 (a,b) AND 5 (a,b) ABOUT HERE

Analysis of *binyan* verb patterns across verb types (lemmas) reveals highly similar distributions in the two corpora, reflecting the patterning of the Hebrew lexicon into three clusters, as follows. One third of the verb types in both spoken and written input consists of the basic *qal* pattern; one third consists of the highly transitive *hif'il* and the currently basic, denominal and transitive *pi'el*; and one third consists of middle-voice, low-transitivity *hitpa'el* and *nif'al*. The spoken sample includes no verbs in the two strictly passive *binyan* patterns, *pu'al* and *huf'al,* while only some 4% of the verb types in the written sample are in passive forms. [11]. Verb tokens, in contrast, reflect the different patterning of lexicons in two distinct communicative settings, targeting different age groups. The spoken parental input addressed to toddlers consists of mainly transitive verbs, with almost three-quarters in *qal*, 15% *hif'il*, and 10% *pi'el*, and only 5% of lower transitivity *hitpa'el* plus *nif'al*. These distributions are also skewed towards the older, less regular system of *qal*, *hif'il*, and *nif'al*, and less towards the 'verb minting machine' of *pi'el* and *hitpa'el* in current Hebrew (Nir 1993). Tokens in the written input, while still dominated by *qal* (60%) and *hif'il* and *pi'el* (25%), include relatively more low-transitivity *hitpa'el* plus *nif'al* forms (11%), they are more balanced across the two *binyan* sub-systems, and have 1% passive verbs.

Morphological families in frequent verbs. The distribution of root-related families in input to young children is clearly demonstrated in Table 4 above, showing the 40 most frequent verbs in the two corpora. The overwhelming majority of verbs on both lists – 85% of the spoken verbs and 95% of the written verbs – are unique in the sense that they do not share roots with any of the other verbs among the 40, showing a breakdown of one verb=one root. The spoken list of 40 frequent verbs has three root-related families, all consisting of two verbs: (1) *ba* 'come' and *hevi* 'bring' (root BWA), ranking 1st and 5th respectively on the list; (2) *ra'a* 'see' and *her'a* 'show' (root RAY), ranking 3rd and 19th respectively; and (3) *hoci* 'take out' and *yaca* 'go out' (root YCA), ranking 16th and 25th respectively. All three pairs express basic active/causative relations alternating *qal* with *hif'il*, and all three are based on defective roots. In fact, each pair shares a single consonant (*b/v*, *r* and *c* respectively), making it extremely hard to construe them as families sharing structural skeletons. The written input contains a single root-related pair - *ba* ‘come’ and *hevi* ‘bring’ (related via BWA), ranking 7th and 23rd respectively on the list. In other words, the 40 most frequent verbs that children are exposed to do not display root-related relationships, with causativity the most basic semantic relationship they express.

Morphological families in the two corpora

Our final analysis relates to the distribution of root-related morphological families across the entire data-base. Figure 6 presents the distribution of verb families sharing the same root across two, three, and four verbs with different *binyan* patterns, as well as roots that occur in only one *binyan.*

PLEASE INSERT FIGURE 6 ABOUT HERE

Single binyan verbs. The distributions of root-based families in the two input corpora are similar: most of the verbs -- 72% in the spoken corpus and 69% in the written respectively – are not root-related. A quarter of the spoken verbs and 23% of the written verbs reside in two-*binyan* families. Larger families with three members occupy 3% of the spoken verbs and 8% of the written, and a negligible number (one in speech, five written) have four members. Thus, the overwhelming majority of the verbs children encounter is either root singletons or else they share a root with one verb in another *binyan*. If these verbs participate in large root-based families in later, more advanced lexicons, then those families are not yet present in this preschool childhood input. A tentative conclusion is that in early childhood, lexical verb learning is verb learning, and consequently – that roots initially ride on verbs. But at the same time, this does not mean that young Hebrew-speaking children do not experience the internal morphological structure of verbs. Recall that roots and patterns are already present in the temporal (and inflected) stems of each *binyan.* Since input contains examples of all five non-passive *binyamin*, children are exposed to their temporal stems, which are constructed of the same root in different, *binyan-*

specific temporal patterns, as in the infinitive / past tense alternations of the root ZHR in *le-hizaher* 'to-take-care' */ nizhárti* 'took-care:1st,Sg = 'I took care'. This means that early one root / one *binyan* input can already be construed as having the internal root-and-pattern structures that will later on permeate the verb lexicon.

Two-binyan verb families. Our data indicate that the notion of the root as relating different verb lemmas 'starts small' in two-*binyan* verb families, which carry the load of early learning about verb morphological derivation in Hebrew. Such reduced entropy (that is, less morphological complexity) at the beginning of the learning process is evidenced by the fact that only 30% of the verbs in both corpora are related via roots, and by the fact that the overwhelming majority of root-related families involve only two *binyamin*.

Two more factors facilitate early root learning. One is the fact that of the two members of the family, one lemma (usually the more basic) always makes a much larger, more prominent appearance than the other. For example, basic *axal* 'eat' (*qal*) in the spoken corpus has 1,252 tokens, while causative *he'exil* 'feed' (*hif'il*) from the same root has only 11. In the same corpus, transitive *gilgel* 'roll' (*pi'el*) has 23 tokens, while intransitive, inchoative *hitgalgel* (*hitpa'el*) denoting self-locomotion has 46 tokens. In the same way, basic *yaca* 'leave, go out' (*qal*) in the written corpus has 137 tokens, while causative *hoci* 'take out' (*hif'il*) has only 25; and *zaxal* 'crawl' (*qal*) has 11 tokens, while reflexive *hizdaxel* 'drag oneself' (*hitpa'el)* makes only one appearance. Thus, 'starting small' relates to relying mostly on one member as a stepping stone to a root-based family which entrenches basic root meaning by frequent repetition, while a second, root-related verb with a more complex meaning modulation makes a smaller appearance, making it easy to relate the two. The side benefit is also learning *binyan* pairs, and the transitivity values these pairs convey, by frequent association.

There is yet another side to starting small – the fact that input relies on a small number of major two-*binyan,* same-root alternations. The spoken input has three pairs of same-root alternations, all within the same *binyan* sub-system. (1) 30 *qal / hif'il* basic / causative sets such as *axal / he'exil* 'eat / feed' (root AKL), *yada / hodía* 'know / let know' (root YDI), and *ala-he'ela* 'go up – bring up' (root ILY); (2) 21 *qal / nif'al* basic / middle ~ inchoative sets such as *patax / niftax* 'open,Trans / open up' (root PTĦ), or *šavar / nišbar* 'break, Trans / break, Intr' (root ŠBR); and (3) 37 *pi'el / hitpa'el* basic / middle ~ inchoative sets such as *lixlex / hitlaxlex* 'dirty / get dirty' (root LKLK), or *siyem / histayem* 'finish / end' (root SYM). The written input contains four pairs of same-root alternations, three within the same *binyan* sub-system and one linking the two subsystems: (1) 37 *qal / hif'il* basic / causative alternations such as *yaca / hoci* 'leave, go out – take out' (root YCA); (2) 28 *qal / nif'al* basic / middle ~ inchoative alternations such as *daxaf / nidxaf* 'push,Trans / be-pushed' (root DĦP), or *paga / nifga* 'hurt / get hurt' (root PGI); (3) 30 *pi'el / hitpa'el* basic / middle ~ inchoative alternations such as *bišel / hitbašel* 'cook / get-cooked' (root BŠL) or *sovev / histovev* 'turn / turn around' (root SBB); and (4) 37 *qal / hitpa'el* basic / middle, e.g., *xafar / hitxaper* 'dig / dig oneself in' (root ĦPR), and *našaf / hitnašef* 'blow / breathe heavily' (root NŠP).

What these numbers indicate is that children are initially exposed to a small set of same-subsystem *binyan* pairs sharing the same root, with a core of an even smaller set of frequent root-*binyan* pairs. Early input thus delivers two small-scale, inter-dependent messages to Hebrew-learning children – one introducing them to the system of root-pattern morphology, while at the same time clearly relating such pairs to enhanced and reduced transitivity values (Berman 1993a,b). Obviously children start verb learning from verbs rather than from roots, as abstract unpronounceable entities that emerge from temporal and derivational verb learning. As predicted, root tokens in early input are largely defective, but the same input helps forge transparent semantic links between pairs of root-related verbs, making it easy to construe the root as the semantic core of the pair.

Three- and four-binyan verb families. The smaller number of families where a single root skeleton relates verbs in three or more different *binyanim* indicates how children may start going beyond the small, semantically transparent basic – middle~inchoative / causative sets. The spoken input includes examples such as *yašav / hošiv / hityašev* 'sit / seat / sit down' (root YŠB) – the basic position verb, causative, inchoative respectively, also adding the iterative *hitpa'el* modulation to the basic/causative pair in the threesome *af / he'if / hit'ofef* 'fly / make- fly / fly away (root IWP); or *kafac / hikpic / kipec* 'jump / make jump / skip' (root QPC), adding the iterative *pi'el* modulation to the canonical basic / causative relations (Berman 1993 a,b). Here, too, the canonical member of the family always carries the major quantitative load, as in *maca / nimca / himci* 'find' (195 occurrences) / 'be-found = exist' (18) / 'invent '(4)' (root MCA). The spoken input thus offers the learning child scaffolding in the form of two semantically related alternations, with a further, less consistent or less semantically transparent modulation, often in the other subsystem. The written input contains fewer canonical, more semantically variegated examples of three-*binyan* families, such as mental *zaxar / nizkar / hizkir* 'remember / recall / remind' (root ZKR). In many of these examples, the *qal* verb is not the canonical carrier of the core root meaning, e.g., *xalaf / hexlif / hitxalef* 'go past, elapse / change, Trans. / exchange with' (root ĦLP). In others, children are presented with variations on the same meaning, as in the *qal* and *nif'al* versions of 'get lost' *avad / ne'evad / ibed* 'be lost / get lost / lose' (root ABD).

Beyond semantics, the extended families also expose children to more morpho-phonological phenomena typical of the verbal system. One is stop/spirant alternations (Ravid

1995), which are particularly apparent when crossing from one subsystem into another (e.g., *lavaš / hilbiš / hitlabeš* 'put on, wear / dress,Trans. / dress oneself', root LBŠ); another is spirant metathesis in *hitpa'el,* e.g., *histaper* 'get a haircut' (cf. *hitlabeš* 'get dressed', *hitparek* 'fall apart', *hitmared* 'rebel'). Moreover, with more *binyan* verb patterns in morphologcial families, exposure to variations of defective stems increases, so that children grow more familiar with the different allomorphs that *binyan* temporal stems take in specific defective classes. For example, *qal w*-medial past and present forms constitute identical, monosyllabic forms (*ba* 'come', root BWA, *zaz* 'move', root ZWZ, and *rac* 'run', root RWC), while *hitpolel* is the *hitpa'el* allomorph for both *w*-medial and double roots -- e.g., *hitkonen* 'get ready', root KWN, and *histovev* 'turn around', root SBB.

Larger root-based families are rare across the data-base. The spoken sample has *ra'a / nir'a / her'a / hitra'a* 'see / appear / show / see each other' from the root RAY (where all occurrences of the reciprocal verb are in the prevalent modal future form *nitra'e* 'we'll see each other') -- a family which beautifully demonstrates basic, middle, causative and reciprocal verb functions. However the written four-verb families present even young children with semantic challenges. For example, the motion verb family based on the root IWP 'fly' represents slightly modulated versions of the same meaning, requiring close attention to the pragmatic context in which they appear (e.g., *af / he'if / ofef / hit'ofef* 'fly / fly,Tr ans/ fly around / fly away'); and semantically extended families such as *našak / nišek / hišik / hitnašek* 'kiss, Literary' / 'kiss', Transitive / 'overlap, interface', Metaphorical / kiss each other', Reciprocal.

Summary and Discussion

The goal of the chapter was to present an initial foray into the nature of child-directed Hebrew verb input. To this end, we analyzed verbs in the spoken parental input to toddlers aged 1;8-2;2, coming to some 300,000 words, and in written narrative texts targeted at preschool and early school-age children aged 1-7 years, containing around 50,000 words. The information conveyed here is new in two senses. First, it provides researchers for the first time with details on a variety of morphological constituents of verbs, hence with a picture of the make-up of the verb lexicon in input to Hebrew-speaking preschoolers. Second, it gives an idea of the sources available to young children in learning the challengingly rich derivational features of Hebrew verbs.

Despite the difference in size, the two corpora contain quite similar numbers of verbs and verb-roots, with the written texts having somewhat more of both. Our analyses reveal that these two sources together expose children to over one thousand (1,186) different verbs and under one-thousand (816) different roots, more than half of which are shared by both corpora. Spoken parental input involves heavy repetition of verb forms, serving as a platform for the initial learning of roots and verb forms. Written texts, targeting a broader and more advanced age range, makes less use of repetition, relying on verb and root systems that have already consolidated, with new verbs constantly entering the repertoire. The most frequent verbs in both corpora are in the basic *qal*, with defective roots; again across both corpora, root types are mostly full, while root tokens are mostly defective, especially in the spoken input; and the largest classes of defective roots are based on semi-vowels (*y*-final and *w/y*-medial). Combined with the prevalent *qal*, these features of the input verbs provide children with a small number of morpho-phonological sites and behaviors, while repetition of different temporal stems (with different agreement markers) is a further aid to mastering the allomorphy involved. Such defective root forms constitute the core root stock of Hebrew, whereas the more mature verb lexicon consists of primarily full roots, which are more widespread in written sources targeting older children –

analogously to findings in other languages for syntactic structures (Cameron-Faulkner & Noble 2013) and for morphology (Nippold & Sun 2008).

With regard to breakdown of *binyan* paradigms in the input data, the core verb types divided almost equally between basic *qal*, largely transitive *hif'il* and *pi'el*, and mainly intransitive or low-transitivity *hitpa'el* and *nif'al.* Verb tokens were overwhelmingly in *qal*, followed by *hif'il* and *pi'el,* with only a few tokens in *hitpa'el* and *nif'al* (see Berman 1993b for similar results). Again unlike adult Hebrew, spoken input to children consists mostly of the older subsystem of *qal*, *hif'il* and *nif'al*, with *pi'el* and *hitpa'el*, which play a major role in denominal verb–formation, lagging behind. Moreover, most roots in both corpora occur in only a single *binyan* – again, mostly *qal*, with no morphological variation in *binyan* alternation in the input; the rest occur in two-*binyan* families, with high semantic transparency; while the remaining few three- (and only occasional few four-) *binyan* families contain more lexically specific, metaphorical, or non-semantically related members (often with defective roots). This state of affairs yields the hypothesis that, when morphological families grow larger and gain more members in later development, they will also grow both semantically less coherent – and structurally more transparent -- than the small core semi-families represented here. By that time, children will already have gained command not only of the structures and functions of the verb system, but also of the orthography that represents them, a major aid in the consolidation of Hebrew morphology (Ravid 2012).

The road to root learning in Hebrew verbs: a tentative model

The native Hebrew learner's task is to eventually construe the root as a structural and semantic morphological core systematically relating verb (and later on, nominal) families. However, as we saw, roots are not a given in early input but rather emerge from the verb input. As we hypothesized, the picture of verbs in input to children emerging from our study is a far cry from the system (re)presented when discussing Hebrew verbs in formal metalinguistic contexts. What toddlers and children mostly hear is different, non-root related verbs, with little evidence of the critical morphological systematicity of roots and *binyanim* hailed as the hallmark of Semitic typology. The system of interrelated roots and patterns that they need to acquire is thus based on verb forms in the input that apparently do not contribute much to its construal. We propose that the path to root learning highlighted by the preceding analyses of verbs in child-directed Hebrew input, consists of "starting small" in the sense of Elman (1993): these mostly structurally opaque one root / one *binyan* verb forms might actually be more child-friendly than previously realized. In other words, Hebrew-speaking children are actually exposed to a smaller, structurally restricted, semantically less challenging, yet morphologically consistent slice of the root-and-*binyan* verb pie. They learn roots and patterns from verbs in small lexical quanta, with gradual increments to morphological and semantic complexity. This constitutes the basis for young children to overcome structural allomorphy and opacity via frequent repetition of different verb forms, while at the same time absorbing more structurally regular root types and learning larger root-based families with increasingly looser semantic ties.

Recall that roots and patterns not only link verbs with different derivational *binyan* conjugations, they also relate different stems inflectionally in the temporal paradigm of each *binyan* (for Present, Past, Future, Imperative, Infinitive). This is important when we consider the nature of the highly-frequent child-directed verb lexicon. The repetitive *qal*-prevalent input provides a safe, restricted habitat within which toddlers can start paying attention to the similar structural skeletons of the *CaCaC* – *CoCeC* -- *PiCCoC* stems (e.g., from the root ZRQ: Past *zarak* 'threw', Present *zorek* 'throws, is-throwing', Future *yizrok* 'will-throw', Imperative *tizrok* 'throw!', Infinitive *lizrok* 'will-throw'). The structures typical of full and frequent defective roots in *qal* can thus be learned without the burden of semantic diversity, since all root skeletons at this point are semantically coherent and transparent, given that they vary within the same verb lemma. What varies is the temporal value of the patterns, which is also presented systematically to toddlers: the shift from modal to present tense and then past tense stems reduces structural opacity and highlights root structure (Ashkenazi 2015; Lustigman 2013a,b). The few frequent non-*qal* verbs provide children with the opportunity to observe and interrelate more, and different, within-*binyan* patterning of a kind less rich than that of *qal* (cf. the consistent stems of *pi'el and hitpa'el* across different Tense/Mood paradigms). Thus, the notion of root structure and semantics can already be gauged from isolated verbs in early acquisition.

Even though most verbs in spoken and written input do not display same-root families, they do support morphological learning. About 800 such verbs in the input enable children to learn the forms of five relevant *binyanim*, first with defective roots, later on increasingly, with full, regular roots. This core verb lexicon is the learning arena of the ten different temporal stems of the Hebrew *binyanim* (three temporal paradigns in *qal*, two each in *nif'al, hif'il* and *pi'el*, and one in *hitpa'el*) and their high-frequency defective allomorphs, which present children with multiply recurring combinations of roots in their respective temporal *binyan*-specific templates. The system emerges by relating clusters of temporal stems with the same lexical meanings into five coherent *binyan* patterns. At the same time, the inherent lexical semantics of verbs and the syntactic contexts in which they occur help children construe the basic transitivity values of the *binyan* system, despite the scarcity of root-related families in their input.

The *binyan* system likewise starts small, with two-*binyan* families. Our findings showed that the input relates a smaller number of *binyan* pairs consistently expressing core transitivity contrasts – basic vs. causative, basic vs. inchoative, middle vs. transitive. The first morphological root-related family ties are forged almost exclusively within the same subsystem, with frequent *binyan* pairs clearly relating forms and transitivity values, highlighting shared, similar root skeletons. Here, too, root semantics is still transparent, with minor transitivity modulations that could aid in initial root construal and subsequent derivational verb morphology.

Productive verb knowledge is initiated by a few three- and four-*binyan* families, forerunners of the richly connected verb system that Hebrew scholars take for granted. These morphological families are often based on new, less frequent, full roots, already demonstrating more complex form-meaning relationships both within and across the two subsystems. Already consolidated knowledge helps older children construe larger, multi-verb families with increasingly loose semantic relations and less prototypical transitivity values. At the same time, lexical specificity increases, with verbs gaining distinct meanings despite structural root relations, and root depth entering the picture in the form of multiple meanings of the same root in the same *binyan* (e.g., *kara*, based on QRA that can mean 'call out', 'call by name', and 'read'). What our study indicates, then, is that the birth of what we recognize as the root-based morphological family can be attributed largely to the contribution of written input.

Future research[12] is necessary to compare these findings to the structure of the verb lexicons in spoken children's productions across toddlerhood, childhood and adolescence and in written monologic texts by schoolgoing children and adolescents. We can assume that young children's spoken productions of verbs, roots and *binyanim* would constitute a subset of the structural and semantic distributions in parental input. However, the question remains open regarding these distributions in later language development, especially in written texts of different genres. As the corpora of written Hebrew under current psycholinguistic analysis were not produced by language experts (Berman & Verhoeven 2002; Berman & Ravid 2009), we might assume that the lemma distributions would continue to adhere to the canonical Hebrew distributions revealed in the current study, perhaps resembling the token distributions in the written texts we analyzed.

TABLE 1. Information about the child-directed corpora used in the current analyses

Ashkenazi Corpus (Ashkenazi 2015) **Spoken Input to Toddlers** 299,461 word tokens	**Ben Zadok & Levie, Grunwald & Bratslavsky Corpora** **Written Input to Preschoolers and Young Schoolchildren** 49,384 word tokens
Dense recordings of two Hebrew-speaking dyads: Boy and parent, girl and parent (two different families) over a period of six months, from 1;8 – 2;2 Recordings took place three times a week, one hour each time, in naturalistic spontaneous interaction with parents. All verbs produced by parents and children transcribed, coded, and morphologically analyzed In the current context we report only on CDS = Child Directed Speech	(i) Ben Zadok & Levie (2014) Morphological analyses of all verbs in 100 storybooks targeting preschoolers aged 1 to 6 years (ii) Grunwald (2014) Morphological analyses of all verbs in 40 school texts for 1st & 2nd graders (6 to 8 year-olds)

TABLE 2. Structural root categories

CATEGORY NAME	ROOT TYPE DESCRIBED	EXAMPLE
Full	Tri-consonantal roots	GDL ‘grow’
Quadriliteral	Quadri-consonantal roots	ŠKNI ‘persuade’
Reduplicated Quadriliteral	Quadri-consonantal roots composed of a reduplicated set	QLQL ‘spoil’
Final Doubled Quadriliteral	Quadri-consonantal roots with the final consonant doubled	IRBB ‘mix’
Denominal Quadriliteral	Quadri-consonantal roots derived from nominals	INYN ‘interest’ from noun inyan ‘interest’
Double	Tri-consonantal roots with two identical final consonants, which create non-canonical and opaque morpho-phonological structures	SBB ‘turn around’
N-initial	Roots with initial N, which deletes in consonant clusters	NPL ‘fall’
Y-initial-C	Roots with initial Y, which deletes in consonant clusters, followed by C	YCB ‘set up’
ʔ(A)-initial	Roots with initial *ʔ*- (A), which shows up as *o* in *qal* modal forms	AHB ‘love’

Y-initial	Roots with initial Y, which shows up as o in consonant clusters	YRD 'get down'
Y/W-medial	Roots with medial Y or W, which create non-canonical and opaque morpho-phonological structures	RYB 'fight' QWM 'get up'
? (A)-final	Roots with final ? (A), which create final open syllables	MCA 'find'
Y-final	Roots with final Y, which create final open syllables	BKY 'cry'
Composites	Roots which belong in more than one defective category	BWA 'come' (W-medial and ? (A)-final)

TABLE 3. Verb and root frequencies in the corpora

	Word tokens	Verb tokens	Verb types	Root types
Spoken parental input to toddlers	299,461	54,810	684	521
Written children's texts	49,384	11,228	1,048	744

TABLE 4a. Forty most frequent verbs in the spoken corpus in descending order of frequency

#	Hebrew verb	Gloss	*binyan*	Frequency
1	*ba*	come	*qal*	5234
2	*raca*	want	*qal*	3876
3	*ra'a*	see	*qal*	3157
4	*asa*	do	*qal*	2949
5	*hevi*	bring	*hif'il*	2469
6	*sam*	put	*qal*	1715
7	*halax*	go	*qal*	1458
8	*amar*	say1	*qal*	1256
9	*axal*	eat	*qal*	1252
10	*yašav*	sit	*qal*	1166
11	*le-hagid*	say2	*hif'il*	1102
12	*yada*	know	*qal*	971
13	*lakax*	take	*qal*	853
14	*kara*	read	*qal*	768
15	*natan*	give	*qal*	705
16	*hoci*	take out	*hif'il*	664
17	*yašan*	sleep	*qal*	651
18	*yaxol*	be able	*qal*	526
19	*her'a*	show	*qal*	519
20	*nafal*	fall	*qal*	517
21	*histakel*	look	*hitpa'el*	511
22	*kara*	happen	*qal*	503
23	*ahav*	love	*qal*	497
24	*azar*	help	*qal*	463
25	*yaca*	go out	*qal*	454
26	*ciyer*	draw	*pi'el*	432
27	*patax*	open	*qal*	420
28	*shar*	sing	*qal*	394
29	*šata*	drink	*qal*	343
30	*sagar*	close	*qal*	342
31	*xika*	wait	*pi'el*	332
32	*siper*	tell	*pi'el*	329
33	*nizhar*	take care	*nif'al*	328
34	*baxa*	cry	*qal*	325
35	*sixek*	play	*pi'el*	318
36	*xipes*	search	*pi'el*	299
37	*ka'av*	hurt	*qal*	295
38	*hexzik*	hold	*hif'il*	232
39	*diber*	talk	*pi'el*	230
40	*sider*	arrange	*pi'el*	226

TABLE 4b. Forty most frequent verbs in the written corpus in descending order of frequency

#	Hebrew verb	Gloss	*binyan*	Frequency
1	*amar*	say1	*qal*	717
2	*ra'a*	see	*qal*	289
3	*halax*	go	*qal*	251
4	*yašav*	sit	*qal*	206
5	*yaxol*	be able	*qal*	204
6	*raca*	want	*qal*	203
7	*ba*	come	*qal*	202
8	*yada*	know	*qal*	192
9	*asa*	do	*qal*	186
10	*kara*	read	*qal*	186
11	*ša'al*	ask	*qal*	185
12	*xašav*	think	*qal*	139
13	*yaca*	go out	*qal*	137
14	*higía*	arrive	*hif'il*	127
15	*axal*	eat	*qal*	122
16	*natan*	give	*qal*	113
17	*ahav*	love	*qal*	110
18	*yašan*	sleep	*qal*	96
19	*rac*	run	*qal*	96
20	*maca*	find	*qal*	92
21	*hitxil*	begin	*hif'il*	89
22	*lakax*	take	*qal*	88
23	*hevi*	bring	*hif'il*	87
24	*amad*	stand	*qal*	87
25	*baxa*	cry	*qal*	85
26	*sixek*	play	*pi'el*	82
27	*šama*	hear	*qal*	79
28	*hibit*	look2	*hif'il*	77
29	*nixnas*	enter	*nif'al*	76
30	*bikeš*	ask	*pi'el*	73
31	*xazar*	go back	*qal*	71
32	*kara*	happen	*qal*	68
33	*siper*	tell	*pi'el*	67
34	*kafac*	jump	*qal*	66
35	*caxak*	laugh	*qal*	65
36	*histakel*	look1	*hitpa'el*	64
37	*ana*	answer	*qal*	64
38	*šar*	sing	*qal*	62
39	*patax*	open	*qal*	58
40	*ca'ak*	yell	*qal*	58

FIGURE 1a. Type distributions of the 10 structural root categories in spoken parental input

N = 521 root types

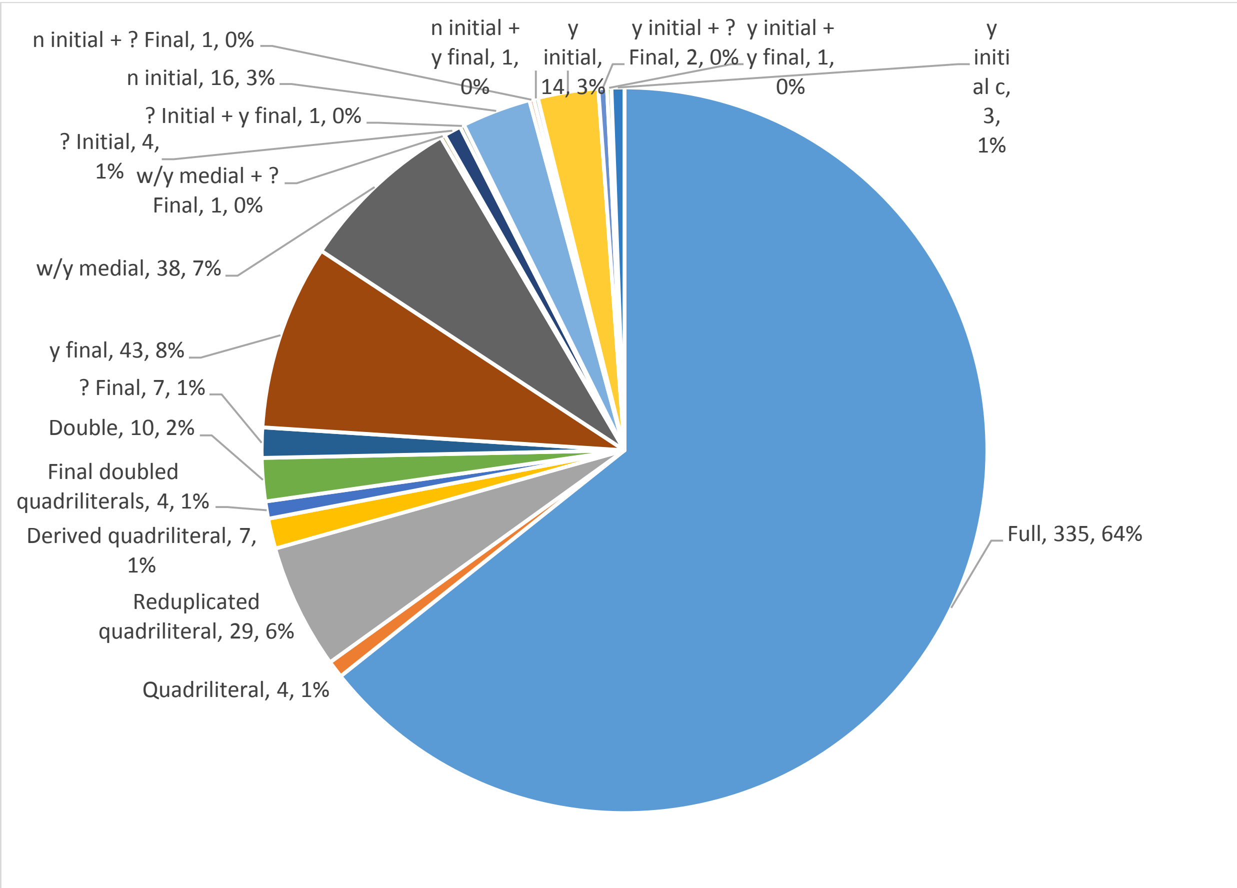


FIGURE 1b. Token distributions of the 10 structural root categories in spoken parental input

N=54,810 root tokens

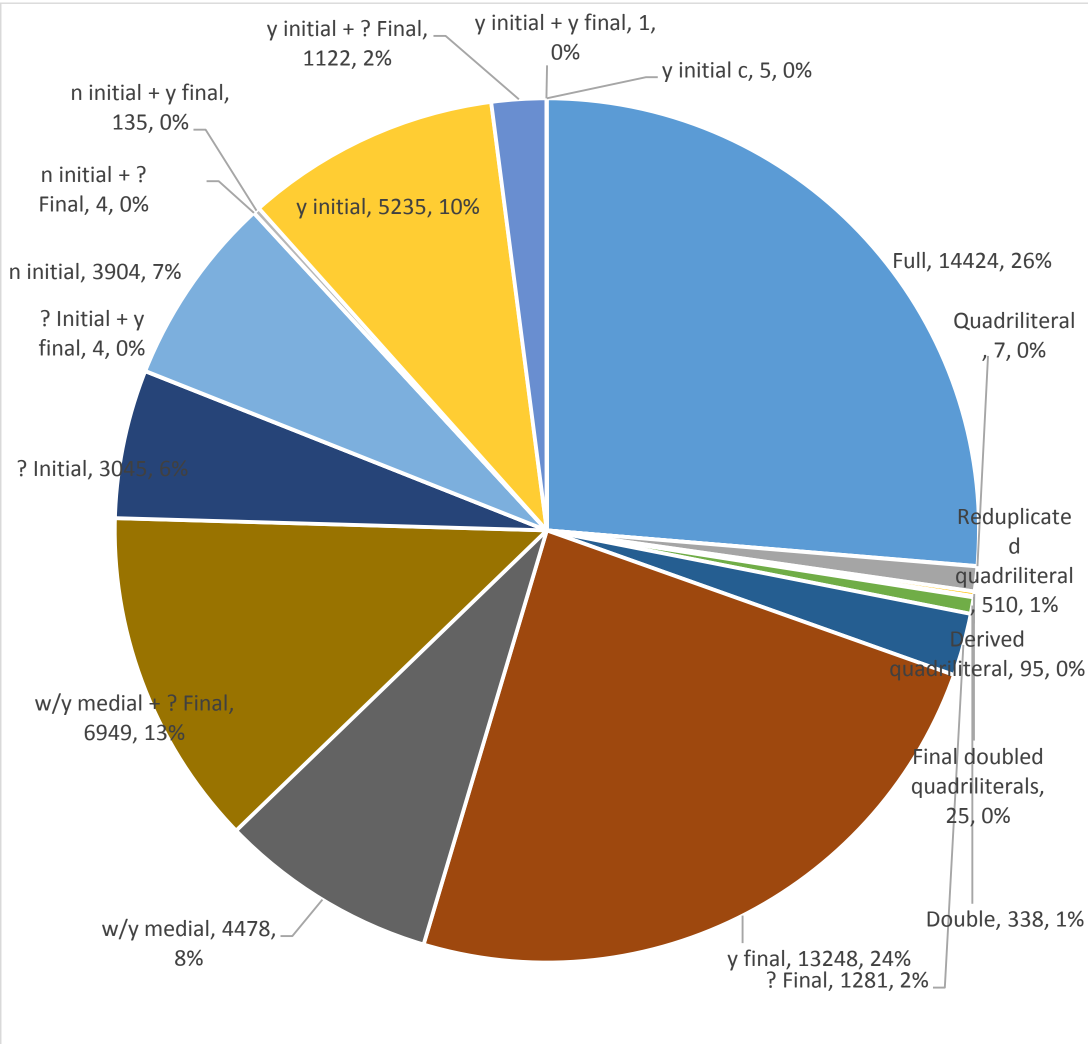


FIGURE 2a. Type distributions of the 10 structural root categories in written children's texts

N = 744 root types

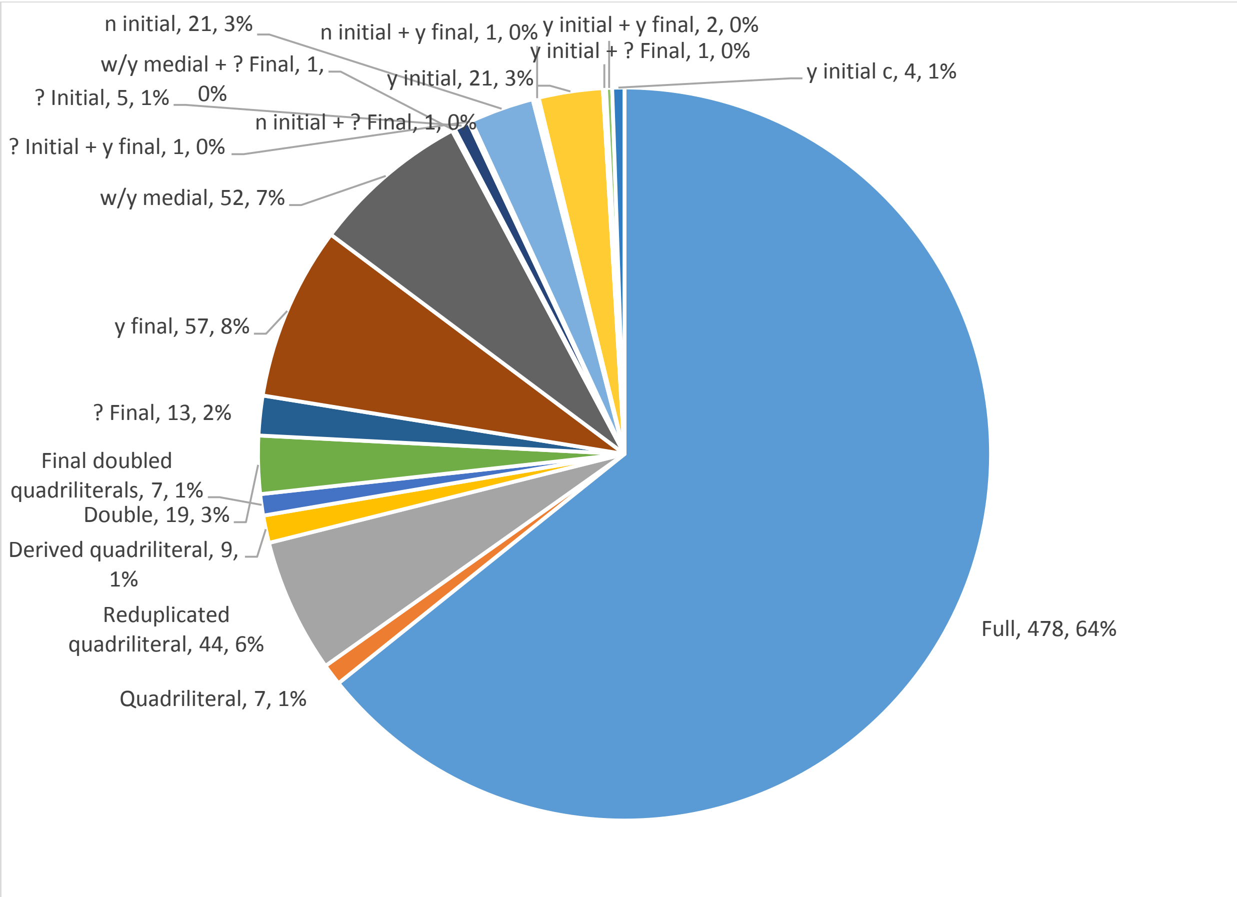


FIGURE 2b. Token distributions of the 10 structural root categories in written children's texts

N = 11,228 root tokens

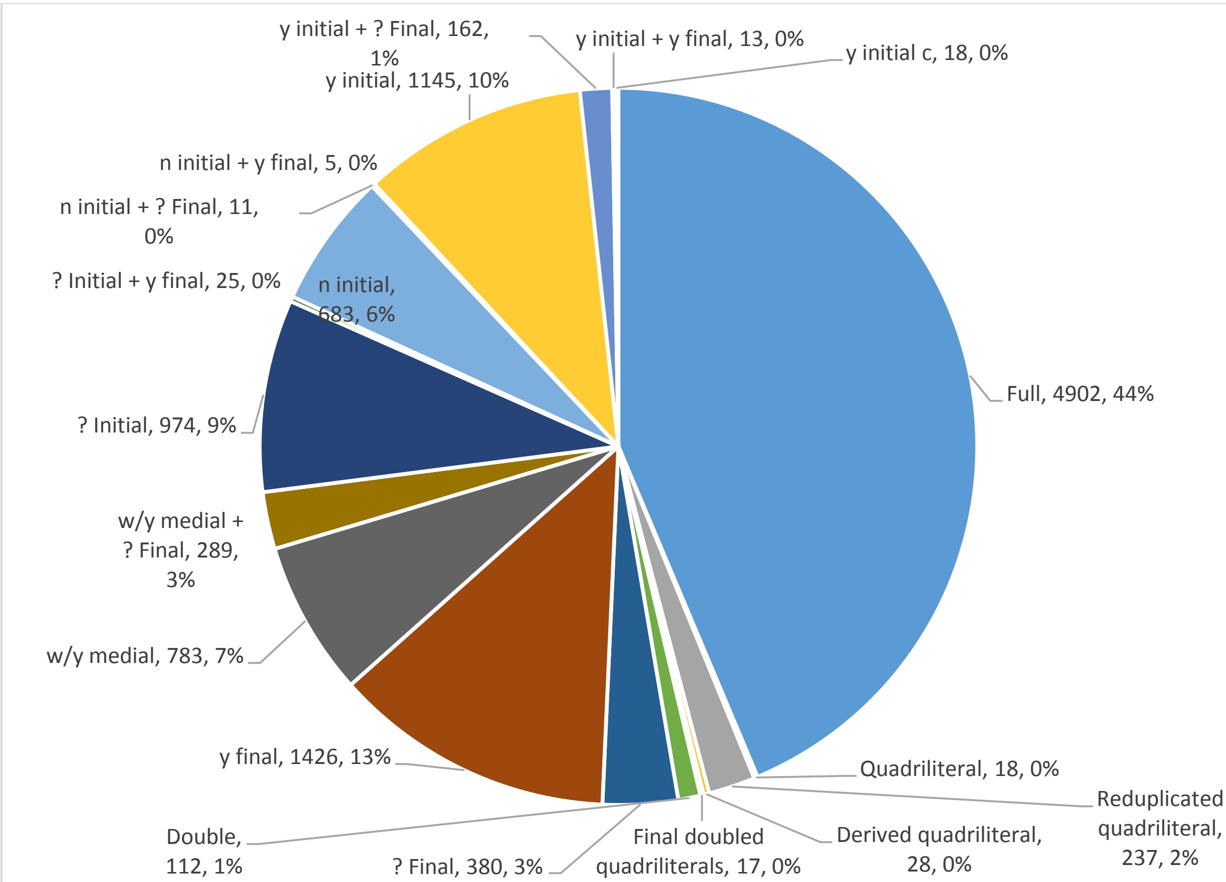


FIGURE 3. Cumulative full (regular) and defective (irregular) root types in parental input to children 1;8-2;2 (Ashkenazi 2015)

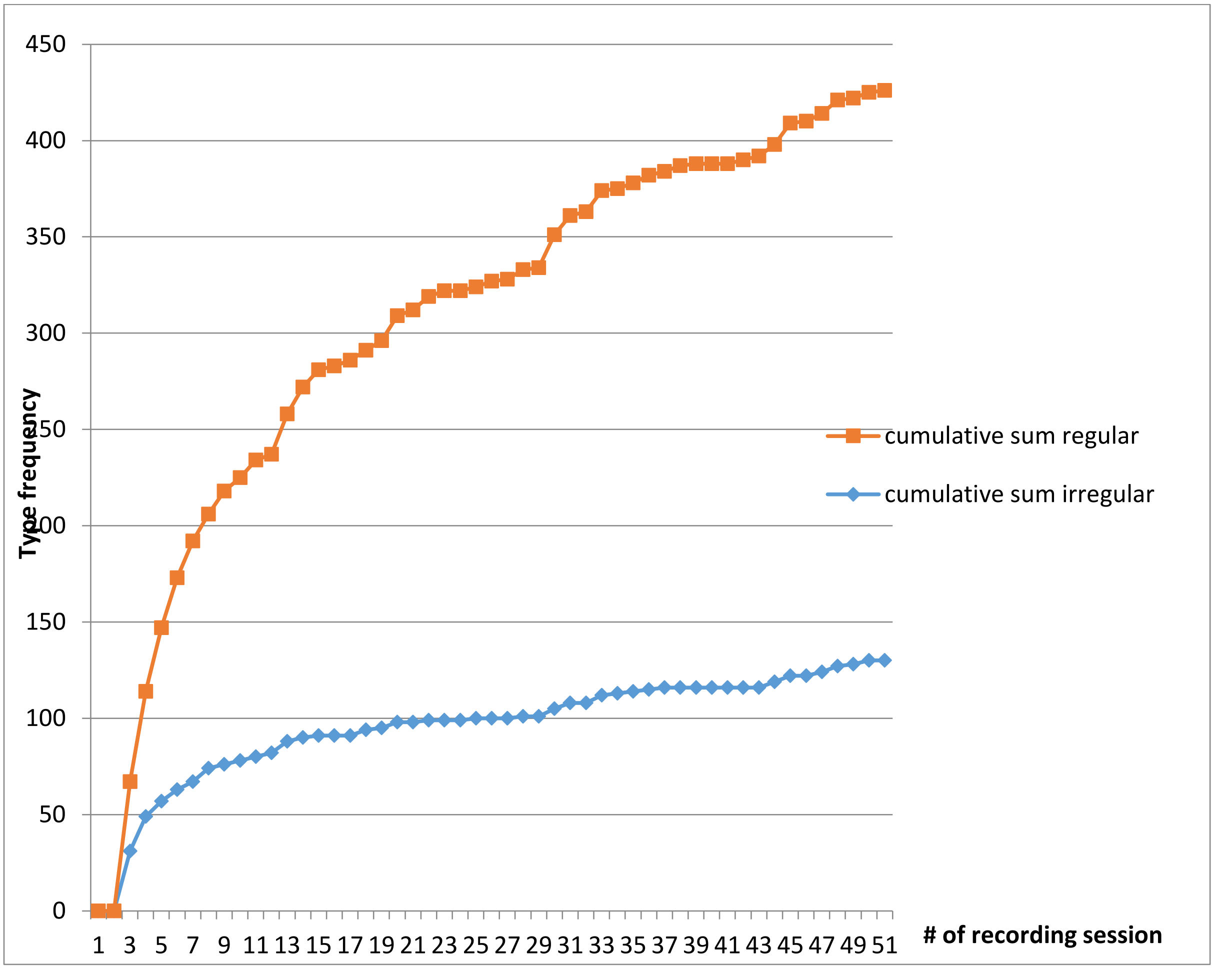


FIGURE 4a. Distributions of the seven *binyan* patterns in 684 verb types, spoken parental input

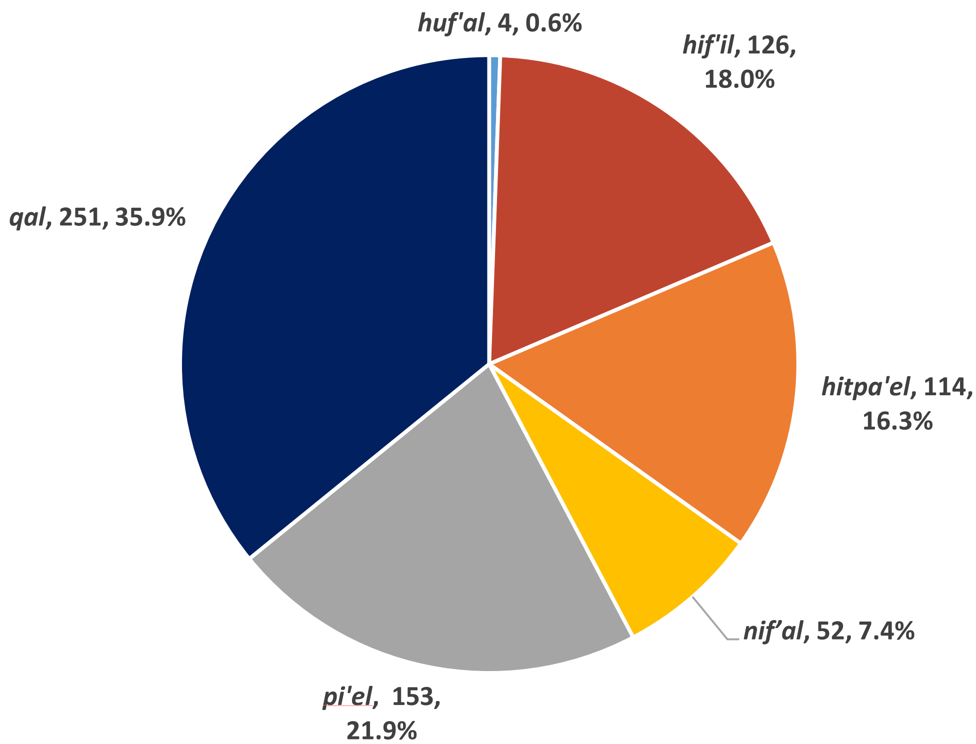


FIGURE 4b. Distributions of the seven *binyan* patterns in 54,810 verb tokens, spoken parental input

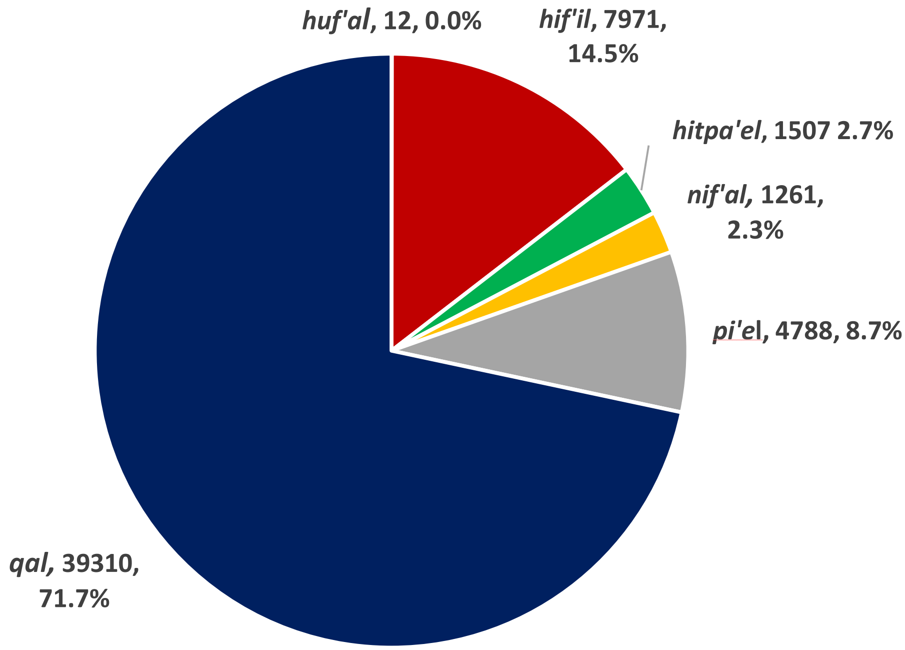


FIGURE 5a. Distributions of the seven *binyan* patterns in 1048 verb types, written children's texts

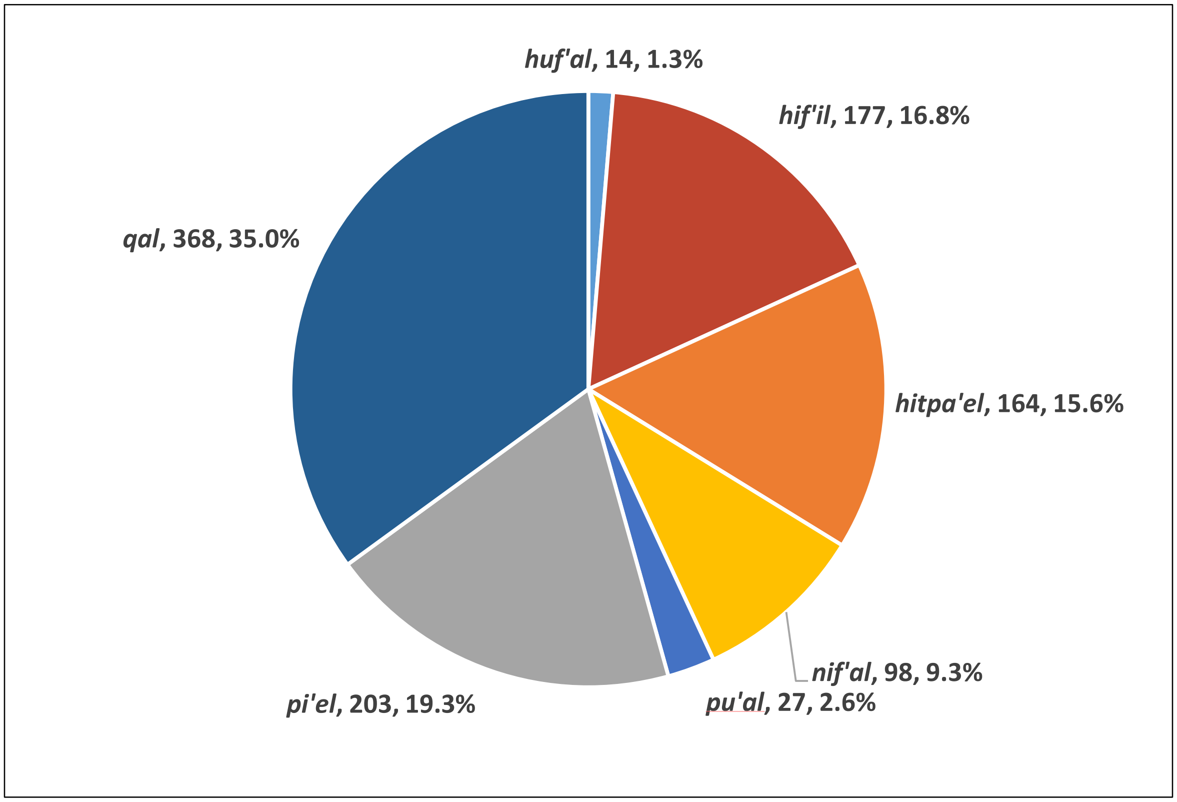


FIGURE 5b. Distributions of the seven *binyan* patterns in 11,228 verb tokens, written children's texts

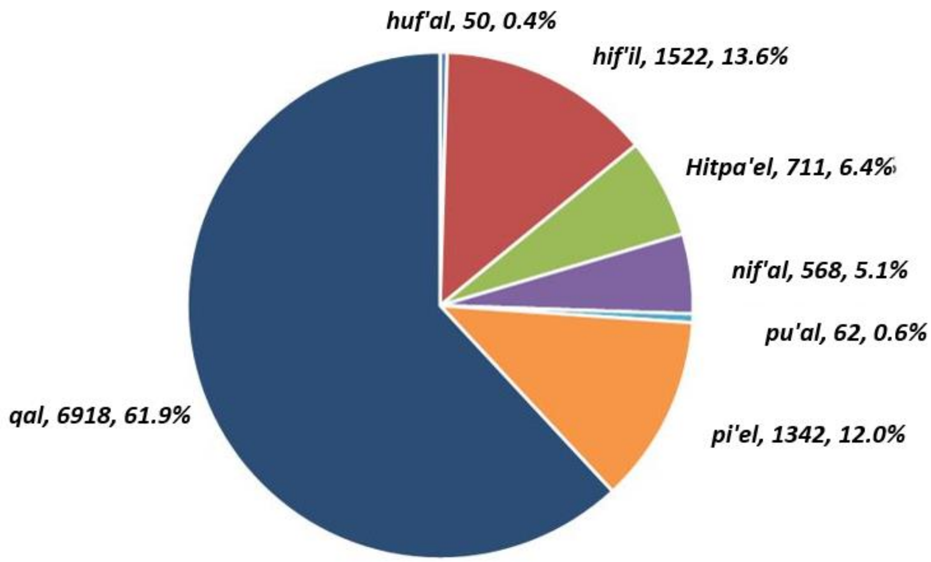
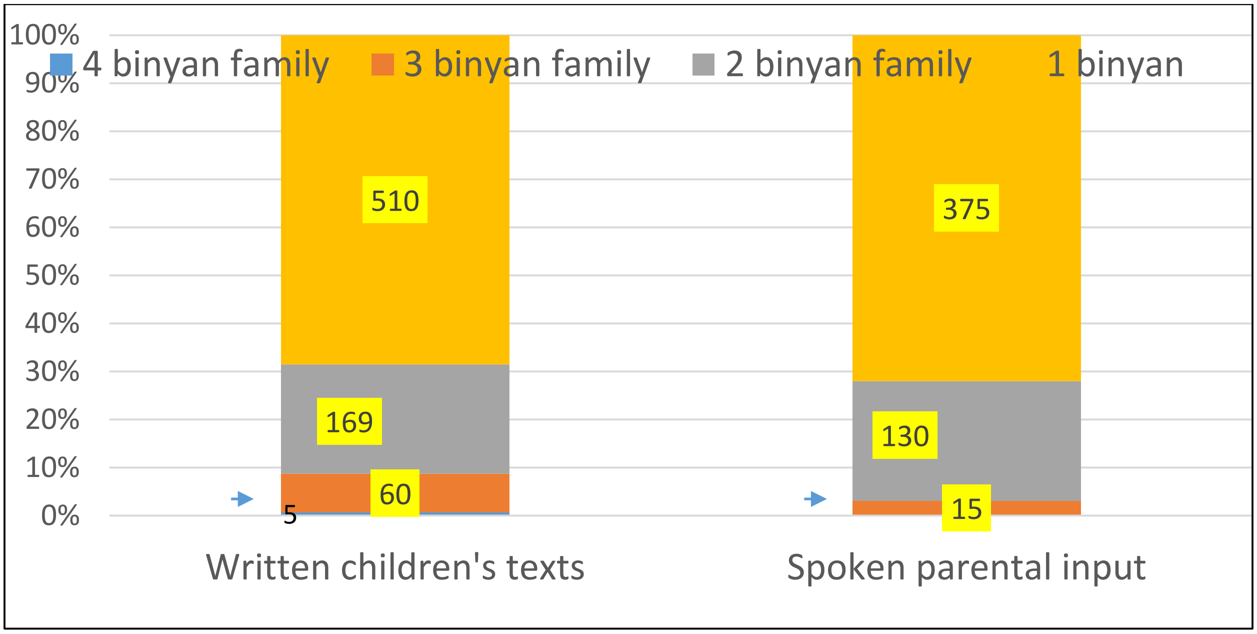


FIGURE 6. Distribution of root-related morphological families in the spoken parental input (N=521 verbs) and written children's texts (N=744 verbs)



References

Acha, J., & Perea, M. 2008. The effect of neighborhood frequency in reading: Evidence with transposed-letter neighbors. *Cognition* 108: 290-300.

Ackerman, F. & Malouf, R. 2013. Morphological organization: The low conditional entropy conjecture. *Language* 89: 429-464.

Armon-Lotem, S. & Berman, R. A. 2003. The emergence of grammar: Early verbs and beyond. *Journal of Child Language* 30: 845-878.

Arnon, I. & Clark, E. V. 2011. Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development* 7: 107-129.

Ashkenazi, O. 2015. Input-output relations in the early acquisition of Hebrew verbs. Unpublished doctoral dissertation, Tel Aviv University.

Ashkenazi, O., Ravid, D. & Gillis, S. Submitted. Input-output relations in the early acquisition of Hebrew verbs: Roots and temporal stems.

Avineri, I. 1976. *The Palace of Patterns*. Tel Aviv: Izre'el [in Hebrew].

Avneyon, E. Ed. 2007. *The New Sapir Dictionary*. Tel Aviv: Itav.

Baayen, R.H. 2010. The directed compound graph of English. An exploration of lexical connectivity and its processing consequences. In S. Olson (ed.), *New Impulses in Word-Formation*, 383-402. Hamburg: Buske Verlag.

Bar On, A. & Ravid, D. 2011. Morphological decoding in Hebrew pseudowords: a developmental study. *Applied Psycholinguistics* 32: 553–581.

Behrens, H. 2006. The input-output relationship in first language acquisition. *Language and Cognitive Processes* 21: 2-24.

Ben Zadok, G. & Levie, R. 2014. Analyses of verb roots in children’s storybooks. Tel Aviv University ms.

Ben Zvi, G. & Levie, R. This volume. Development of Hebrew derivational morphology from preschool to adolescence.

Berman, R.A. 1985. *Acquisition of Hebrew*. Hillsdale, NJ: Erlbaum.

Berman, R.A. 1987. Productivity in the lexicon: New-word formation in Modern Hebrew. *Folia Linguistica* 21: 425-461.

Berman, R.A. 1993a. Marking of verb transitivity by Hebrew-speaking children. *Journal of Child Language* 20: 641-669.

Berman, R.A. 1993b. Developmental perspectives on transitivity: A confluence of cues. In Y. Levy (ed.), *Other Children, Other Languages: Issues in the Theory of Acquisition,* 189-241. Hillsdale, NJ: Erlbaum.

Berman, R.A. 2000. Children's innovative verbs vs. nouns: Structured elicitations and spontaneous coinages. In L. Menn & N. Bernstein-Ratner (eds.), *Methods for Studying Language Production*, 69-93. Mahwah, NJ: Erlbaum.

Berman, R.A. 2003. Children's lexical innovations: developmental perspectives on Hebrew verb-structure. In J. Shimron (ed.), *Language Processing and Acquisition in Languages of Semitic, Root-Based Morphology*, 243-291. Amsterdam: Benjamins.

Berman, R.A. 2012. Revisiting roots in Hebrew: a multi-faceted view. In M. Muchnik and T. Sadan (Eds.), *Studies on Modern Hebrew and Jewish Languages in Honor of Ora (Rodriguez) Schwarzwald*, 132-158. Jerusalem: Carmel Press.

Berman, R.A. in press. Acquiring and expressing temporality in Hebrew: A T/(M/A) Language". *SKASE Journal of Theoretical Linguistics.*

Berman, R.A. & Ravid, D. 2009. Becoming a literate language user: Oral and written text construction across adolescence. In D.R. Olson and N. Torrance (eds), *Cambridge Handbook of Literacy*, 92-111. Cambridge: Cambridge University Press.

Berman, R. A. & Sagi, Y. 1981. Children's word-formation and lexical innovations. *Hebrew Computational Linguistics Bulletin* 18: 31-62 [in Hebrew].

Berman, R.A. & Verhoeven, L., (Editors). 2002. Cross-linguistic perspectives on the development of text production abilities in speech and writing. *Written Language and Literacy*, Volume 5, Parts 1 and 2 (special issue).

Biber, D. 2009. Are there linguistic consequences of literacy? Comparing the potentials of language use in speech and writing. In D.R. Olson & N. Torrance (eds.), *The Cambridge Handbook of Literacy*, 75-91. Cambridge: Cambridge University Press.

Bolozky, S. 1999. *Measuring Productivity in Word Formation*. Leiden: Brill.

Bolozky, S. 2008. *501 Hebrew Verbs*, 2nd edition. Hauppauge: Barron's Educational Series.

Borodkin, K., & Faust, M. 2012. Word retrieval in developmental language impairments: Application of the Tip-of-the-Tongue paradigm. In M. Faust (ed.), *The Handbook of the Neuropsychology of Language*, 963-982. Oxford, UK: Willey-Blackwell.

Brysbaert, M. & Ghyselinck, M. 2006. The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition 13*: 992-1011.

Bybee, J. 2006. From usage to grammar: the mind's response to repetition. *Language* 82: 711-733.

Cameron-Faulkner, T. & Noble, N. 2013. A comparison of book text and Child Directed Speech. *First Language* 33: 268–279.

Clark, E.V. 2003. *The Lexicon in Acquisition.* Cambridge: Cambridge University Press.

Clark, H.H., 1996. *Using Language.* Cambridge: Cambridge University Press.

Clark, E.V. & Clark, H.H. 1979. When nouns surface as verbs. *Language* 55: 767-811.

Davies, M. 2009. The 385+ Million Word Corpus of Contemporary American English (1990-present). *International Journal of Corpus Linguistics* 14: 159-90.

Deutsch, A. & Meir, A. 2011. The role of the root morpheme in mediating word production in Hebrew. *Language and Cognitive Processes* 26: 716-744.

Douani, O. 2014. Register judgments of Hebrew adjectives. MA thesis, the Department of Communication Disorders, Tel Aviv University.

Elman, J.L. 1993. Learning and development in neural networks: the importance of starting small. *Cognition* 48: 71-99.

Even-Shoshan, A. 2003. *The New Even Shoshan Hebrew Dictionary*. Tel Aviv: Encyclopedias.

Frishkoff, G.A., Collins-Thompson, K., Perfetti, C.A. & Callan, J. 2008. Measuring incremental changes in word knowledge: Experimental validation and implications for learning and assessment. *Behavior Research Methods* 40: 907-925.

Fromkin, V. (ed.) 1980. *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. San Francisco: Academic Press.

Frost, R. 2012. Towards a universal model of reading. Target article. *Behavioral and Brain Sciences* 35: 263- 279.

Frost, R., Deutsch, A. & Forster, K.I. 2000. Decomposing morphologically complex words in a nonlinear morphology. *Journal of Experimental Psychology: Learning, Memory and Cognition* 26: 751-765.

Gillis, S. & Ravid, D. 2006. Typological effects on spelling development: a crosslinguistic study of Hebrew and Dutch. *Journal of Child Language 33*: 621-659.

Givón, T. 2009. *The Genesis of Syntactic Complexity: Diachrony, Ontogeny, Neuro-Cognition, Evolution*. Amsterdam: John Benjamins.

Goldberg, A. E., Casenhiser, D. & White, T. R. 2007. Constructions as categories of language, *New Ideas in Psychology* 10: 91-92.

Graves, W. W., Desai, R., Humphreys, C., Seidenberg, M. S., & Binder, J. R. 2010. Neural systems for reading aloud: A multiparametric approach. *Cerebral Cortex* 20: 1799–1815.Grunwald, T. 2014. Verb roots in young readers' texts. MA thesis, School of Education, Tel Aviv University.

Haspelmath, M & Sims, A. 2010. *Understanding Morphology* (2nd edition). New York: Oxford University Press.

Hoff-Ginsberg, E. 1985. Some contribution of mothers' speech to their children's syntactic growth. *Journal of Child Language* 12: 367-385.

Jaeger, J.J. 2004. *Kid's Slips: What Young Children's Slips of the Tongue Reveal about Language Development.* New York: Psychology Press.

Johnston, R.A. & Barry, C. 2006. Age of acquisition and lexical processing. *Visual Cognition* 13: 789-845.

Juhasz, B.J. 2005. Age-of-acquisition effects in word and picture identification. *Psychological Bulletin 131*: 684-712.

Kamalski, J., Lentz, L., Sanders, T. & Zwaan, R.A. (2008). The forewarning effect of coherence markers in persuasive discourse: Evidence from persuasion and processing. *Discourse Processes*, 45: 545-579.

Katz, L., Brancazio, L., Irwin, J., Katz, S., Magnuson, J. & Whalen, D.H. 2012. What lexical decision and naming tell us about reading. *Reading and Writing* 25: 1259-1282.

Levie, R., Ben Zvi, G. & Ravid, D. Submitted. Morpho-lexical development in language-impaired and typically developing Hebrew-speaking children from two SES backgrounds.

Lieven, E. V. M., Pine, J. M. & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187–219.

Lustigman, L. 2013a. Developing structural specification: Productivity in early Hebrew verb usage. *First Language* 33: 47–67.

Lustigman, L. 2013b. Linguistic interfaces in early acquisition: lexical classes and grammatical systems in Hebrew child language. Unpublished doctoral dissertation, Tel Aviv University.

Malsen, R., Theakston, A., Lieven, E. & Tomasello, M. 2004. A dense corpus study of past tense and plural over-regularization in English. *Journal of Speech, Language and Hearing Research* 47: 1319-1333.

Mariscal, S. 2009. Early acquisition of gender agreement in the Spanish noun phrase: starting small. *Journal of Child Language* 36: 143 – 171.

Marslen-Wilson, W.D. 2007. Morphological processes in language comprehension. In G. Gaskell (ed.), *Oxford Handbook of Psycholinguistics*, 175-193. Oxford: Oxford University Press.

Mason, R. A., & Just, M. 2007. Lexical ambiguity in sentence comprehension. *Brain Research* 1146: 115-127.

Nippold, M.A. & Sun, L. 2008. Knowledge of morphologically complex words: a developmental study of older children and young adolescents. *Language Speech and Hearing Services in School* 39: 365-373.

Nir, R. 1993. *Word Formation in Modern Hebrew*. Tel Aviv: Open University [in Hebrew].

Ornan, U. 1990. Production of words (morphology) in Hebrew. *Leshonenu Laam*, 40-41, 247-363 [in Hebrew].

Paterson, K.B., Alcocka, A. & Liversedge, S.P. 2011. Morphological priming during reading: Evidence from eye movements. *Language and Cognitive Processes* 26: 600-623.

Pine, J. M., Lieven, E. V. M. & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics* 36: 807–30.

Ravid, D. 1990. Internal structure constraints on new-word formation devices in Modern Hebrew. *Folia Linguistica 24*: 289-346.

Ravid, D. 1995. *Language change in child and adult Hebrew: A psycholinguistic perspective.* New York: Oxford University Press.

Ravid, D. 2001. Learning to spell in Hebrew: Phonological and morphological factors. *Reading and Writing* 14: 459-485.

Ravid, D. 2003. A developmental perspective on root perception in Hebrew and Palestinian Arabic. In Y. Shimron (ed.), *Language Processing and Acquisition in Languages of Semitic, Root-Based Morphology*, 293-319. Amsterdam: Benjamins.

Ravid, D. 2005. Hebrew orthography and literacy. In R.M. Joshi & P.G. Aaron (eds.), *Handbook of Orthography and Literacy*, 339-363. Mahwah, NJ: Erlbaum.

Ravid, D. 2006. Word-level morphology: A psycholinguistic perspective on linear formation in Hebrew nominals. *Morphology* 16: 127-148.

Ravid, D. 2008. The dual binyan system in Hebrew. Talk presented at the Vienna Morphology Meeting. University of Vienna, February, 2008.

Ravid, D. 2011. The emergence of the Hebrew verb category in toddlers: A new psycholinguistic perspective. *Literacy and Language* 3: 131-160. [in Hebrew]

Ravid, D. 2012. *Spelling Morphology: The Psycholinguistics of Hebrew Spelling*. New York: Springer.

Ravid, D. & Avidor, A. 1998. Acquisition of derived nominals in Hebrew: Developmental and linguistic principles. *Journal of Child Language* 25: 229-266.

Ravid, D. & Bar On, A. 2005. Manipulating written Hebrew roots across development: The interface of semantic, phonological and orthographic factors. *Reading and Writing* 18: 231-256.

Ravid, D. & Tolchinsky, L. 2002. Developing linguistic literacy: A comprehensive model. *Journal of Child Language* 29: 419-448.

Ravid, D. & Schiff, R. 2006a. Roots and patterns in Hebrew language development: evidence from written morphological analogies. *Reading and Writing* 19*:* 789-818.

Ravid, D. & R. Schiff. 2006b. Morphological abilities in Hebrew-speaking gradeschoolers from two socio-economic backgrounds: An analogy task. *First Language* 26: 381-402.

Ravid, D., Levie, R. & Avivi-Ben Zvi, G. 2003. Morphological disorders. In L. Verhoeven and Hans van Balkom (eds.), *Classification of Developmental Language Disorders: Theoretical Issues and Clinical Implications*, 235-260. Mahwah, NJ: Erlbaum.

Rose, T., Stevenson, M *&* ,.Whitehead, M. (2002, May). The Reuters Corpus Volume 1 – from yesterday's news to tomorrow's language resources. In Proceedings of LREC - 3rd International Conference on Language Resources and Evaluation, ELRA: Paris, pp. 827-832.

Rowland, C. F. & Pine, J. M. (2000). Subject–auxiliary inversion errors and wh-question acquisition: ‘what children do know?’ *Journal of Child Language*, 27, 157–81.

Rubino, R. & Pine, J. (1998). Subject–verb agreement in Brazilian Portuguese: what low error rates hide. *Journal of Child Language*, 25, 35–60.

Savickienė, I., Kempe, V. & Brooks, P. J. 2009. Acquisition of gender agreement in Lithuanian: Exploring the effect of diminutive usage in an elicited production task. *Journal of Child Language* 36: 477- 494.

Schiff, R. & D. Ravid. 2007. Morphological analogies in Hebrew-speaking university students with dyslexia compared with typically developing gradeschoolers. *Journal of Psycholinguistic Research* 36: 237-253.

Schwarzwald, O.R. 1974. Roots, patterns, and the morpheme structure. *Leshonenu* [Our Language] 38: 131-136 [in Hebrew].

Schwarzwald, O.R. 1981. *Grammar and Reality in the Hebrew Verb*. Ramat Gan: Bar Ilan University Press.

Schwarzwald, O.R. 2000. Verbal roots and their links to nouns. In O.R. Schwarzwald, S. Blum-Kulka & E. Olshtain (eds.), *Raphael Nir Jubilee Book*, 426-438. Carmel: Jerusalem [in Hebrew].

Schwarzwald, O.R. 2002. *Modern Hebrew Morphology*. Tel Aviv: The Open University [in Hebrew].

Schwarzwald, O.R. 2013. Defective verbs (I: 673-678). EHLL: Encyclopedia of Hebrew Language and Linguistics, I-IV volumes, New York: Brill.

Schwarzwald, O.R. & Cohen-Gross, D. 2000. The productive noun patterns in Hebrew. In M. Horvits (ed.), *The Language of Contemporary Press: Mina Efron's Memorial Volume*, 148-161. Even Yehuda: Reches [in Hebrew].

Seroussi, B. 2011. The morphology-semantics interface in the mental lexicon: The case of Hebrew. Unpublished doctoral dissertation, Tel Aviv University.

Theakston, A., Lieven, E., Pine, J. & Rowland, C. 2004. Semantic generality, input frequency and the acquisition of syntax. *Journal of Child Language* 31: 61-99.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, Mass: Harvard University Press.

Turner, J. E., Valentine, T., & Ellis, A. W. 1998. Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision. *Memory & Cognition* 26: 1282–1291.

Velan, H., Frost, R., Deutsch, A., & Plaut, D.C. 2005. The processing of root morphemes in Hebrew: Contrasting localist and distributed accounts. *Language and Cognitive Processes* 20: 169–206.

Vitevitch, M. & Luce, P. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40: 374–408.

Wilson, S. (2003). Lexically specfiic constructions in the acquisition of inflection in English. *Journal of Child Language*, 30, 75-115.

Zevin, J.D. & M. S. Seidenberg. 2002. Age of Acquisition effects in word reading and other tasks. *Journal of Memory and Language* 47: 1–29.

Notes

*The study was supported by Grant No. 285/13 from the Israel Science Foundation (ISF) for the study of "Input-output patterns in the acquisition of Hebrew root usage: A corpus-based psycholinguistic study", Dorit Ravid PI.

---

[1] The root PIL standing for 'act, do' is traditionally used in naming the *binyamin*, and *pa'al* is often termed *qal* literally 'light, not heavy' because of its syllabic structure.

[2] We deliberately use full, non-defective or "strong" verbs for illustration, e.g., GDL 'grow', SRG 'knit'.

[3] Plural of *binyan*.

[4] Explicit linguistic markers that make coherence relations explicit (Kamalski, Lentz, Sanders & Zwaan (2008).

[5] That is, we counted, for example, the historical root KWN meaning 'prepare' and 'direct' respectively as two different roots based on their different structures (medial *w* vs. *v*) and differential morphophonological behavior (compare *hitkonen* 'get ready' and *hitkaven* 'mean', both in the same *binyan*).

[6] In principle, we could count temporal stems within the same *binyan* as types, e.g., *hevi* 'brought', *mevi* 'brings', *yavi* 'will bring', etc. This is not, however, of concern in the present study.

[7] Thus, root tokens may result from the multiple occurrences of a root in the temporal stems of the same *binyan*, for example, the temporal paradigm of the verb *ba* 'come' in Past, Present, Future, etc. They may also reflect a morphological family, that is, occurrences of the same root in (the temporal stems of) different *binyan* patterns, for example, the root BWA occurring in the verbs *ba* 'come' and *hevi* 'bring'.

[8] Repetition included both repetitions of a given verb form (*bói, bói, bói* 'come: Imper, Fem,Sg x 3) or of any one of the 25 possible different verb forms per verb lemma (e.g., *bo:*'come: Imper, Masc, Sg', *bói*: Imper, Fem, Sg', *la-vo:* Inf, *bánu*: Past, 1st, Plur' , *báti* 'Past, 1st, Sg','etc.). The distributions of verb temporal and agreement inflections would of course differ greatly for verbs with different meanings and functions.

[9] A newer study (Raz, in preparation) reports 335 root types in the written corpus of the cross-linguistic project on developing text production, consisting of narratives and expository texts produced by 4th, 7th and 11th graders, compared with university-educated adults (Berman 2008; Berman & Verhoeven 2002). Since about half of these already occur in the current corpus, this might bring the number of roots in educated Hebrew to about 1,000.

[10] Although restricted to the *pi'el-pu'al-hitpa'el* subsystem of *binyan* alternations.

[11] In most cases, these took the form of resultative adjectives in the present-tense participial *benoni* (Berman 1994).

[12] Already under way in our Language and Discourse TAU lab.