Journal of Child Language (2023), 1–27
doi:10.1017/S030500092200068X

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The development of Hebrew zero and pronominal subject realization in the context of first and second person

Elitzur DATTNER and Dorit RAVID

Tel Aviv University
Email: elitzurd@tauex.tau.ac.il

(Received 30 September 2021; revised 12 September 2022; accepted 06 November 2022)

Abstract

The study investigates the acquisition of Hebrew zero and pronominal subjects in the context of first and second person. We provide distributional evidence relative to verb tense, number, person, and conversational utterance type, in a peer-talk corpus (2;0-8;0 years). Findings show that acquisition starts early on, that verb inflectional morphology is crucial for the development of pronominal subjects, and that communicative contexts affect subject realization. Zero and pronominal subjects are not evenly distributed relative to the study variables, and cannot be treated as an alternation. A cluster analysis shows that each realization is linked to a distinguishable usage pattern, corresponding with particular discursive and communicative functions. These are defined as three Discourse Profile Constructions: (A) "calling for action" by 1st.Pl.Fut zero subjects (3;0 year olds); (B) "commenting on the interlocutor's actions" by 2nd.Sg.Past zero subjects (ages 4;0-6;0); and (C) "planning one's own actions" by 1st.Sg.Fut pronominal subjects (7;0-8;0 year olds).

Keywords: Grammatical subject; Pronoun; Null subject; Hebrew; Discourse Profile Constructions

Introduction

Grammatical subjects in subject-prominent languages (Li & Thompson, 1976) can be realized explicitly as either a lexical noun or as a pronoun. Some languages also allow zero reference in the subject argument, mostly known as null subjects (e.g., Berman, 1980; Borer, 1989), relative to discourse-pragmatic contexts and typological attributes such as predicate type and word order (Ariel, 2001; Ravid, 1995). Across languages, gaining command of grammatical subjects in general, and of pronominal and zero subjects in particular, is argued to be related to a complex bundle of factors, including usage patterns typical of the target language (e.g., Alonso-Ovalle, Fernández-Solera, Frazier, & Clifton, 2002; Conwell, 2020; Forsythe, Greeson, & Schmitt, 2019), semantic and discursive factors (e.g., Arnon, 2010; Li & Thompson, 1976), or parameter settings (Hyams, 2011). The current study investigates the acquisition of zero and pronominal subjects

in Hebrew, a language marking verbs in agreement with their grammatical subjects, e.g., *sixákti* 'played.1st.Sg' / *ani sixákti* 'I played.1st.Sg'. Specifically, we ask whether children's path of grammatical subject acquisition can be explained within a Usage Based Construction Grammar framework (Bybee, 2006), as developing a specific type of construction – namely, a Discourse Profile Construction.

The Discourse Profile Construction Hypothesis (Dattner, 2015, 2019; Dattner, Kertes, Zwilling, & Ravid, 2019) shows that communicative functions (e.g., "calling for action", or "commenting on the actions of the other") are probabilistically linked to a set of parameters conveying morphological, syntactic, pragmatic, and conversational information, forming conventional pairings of form and function. These conventions are termed Discourse Profile Constructions as they constitute basic discourse-level constructions, extending the Argument Structure Constructions notion which concerns basic events and clause level constructions (Goldberg, 2003). Discourse Profile Constructions emerge from linguistic and extra-linguistic properties that pattern together in language use, simultaneously serving a particular communicative function. Taking this hypothesis regarding language use as our framework, the present paper's research questions are (i) whether such usage patterns can be found in the use of zero and pronominal subjects in child language in Hebrew, regarding the properties of the subject referent, verb tense, and the role of the utterance in the flow of conversation; and (ii) whether we can detect specific discursive or communicative functions served by these patterns. In order to do so, we examine the development of Hebrew zero and pronominal subjects in a corpus of peer talk interactions of children aged 2;0-8;0 years, focusing on first and second person subjects, in past and future tenses. We argue that these subjects develop not as a set within an alternation, but as belonging to different constructions in the sense of Construction Grammar, and specifically, as belonging to different Discourse Profile Constructions that couple a usage pattern with a communicative function.

Learning to express grammatical subjects

The investigation of syntactic development is interested in two major questions, regarding (i) the time frame in which children produce grammatical constructions in a creative yet conventional manner (Tomasello, 2000); and (ii) the factors that impact this attainment. Together, these queries are involved in explaining how children gain command of grammatical relations in their language, based on the distributional evidence of semantic notions over structures (Morris, Cottrell, & Elman, 2000). A major part of these grammatical relations involves the notion of grammatical subject (Valian, 1991). As grammatical subjects are referential expressions conveying discourse-pragmatic information, children need to learn how to refer to entities in the world, which is inevitably related to children's growing ability to talk about events, activities, and situations (Conwell, 2020; Golinkoff & Hirsh-Pasek, 2008; Serratrice & Allen, 2015). Specifically, children need to learn the subset of subject-referring expressions in their language as they are used in different communicative contexts, and thus relate verbs to their argument structures (Serratrice, Sorace, & Paoli, 2004; Veneziano & Clark, 2016).

Dattner et al. (2019) showed that gaining command of the Hebrew syntactic category of grammatical subjects involves mastering its conventional discourse profiles, i.e., unique patterns of usage. These usage patterns are part of the form-meaning pairings of the ambient language, from which the meanings of linguistic elements emerge (Dattner, 2019). Dattner et al. (2019) found three types of subject usage patterns. First, from age 3;0

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

and up, initiated utterances with a pre-verbal, first person plural pronominal subject and a future tense verb, serving to plan joint actions, e.g., *come, we'll draw a bug.* A second pattern, prevalent in all age groups older than 2;0, pre-verbal 1st and 2nd person singular pronominal subjects, in present and past tense, in both initiated and response utterances. This usage pattern describes and comments on current actions and past events, and is used for narrating events. For example, *I couldn't open it.* Finally, a usage pattern consisting of post-verbal lexical subjects with verbs in the present tense, used to introduce new topics to the discourse and to refer to less accessible entities outside of the conversational arena. For example, *are there any more cookies?* This usage pattern reflects the development of conversation, responding to, repeating, or resonating a previous utterance, as it is used differentially across development.

The current study zooms in on learning to express pronominal subjects in finite-verb, subject-requiring contexts, given the specific conditions and constructions of Hebrew pronominal expression. Our specific research question focuses on children's gaining command of pronominal and zero subject expression in Hebrew, e.g., *ani nixnásti* 'I entered.1st.Sg.', versus *nixnásti* 'entered.1st.Sg.'. We assume that the conditions of nominal realization in all syntactic positions and verb arguments across languages are related to the degree of nominal accessibility in discourse (Ariel, 2001; Du Bois, 1987; Keenan & Comrie, 1977; Serratrice, 2005). Accordingly, the current study examines the spontaneous peer conversation of Hebrew-speaking preschool children to determine the usage patterns of pronominal and zero realizations of nominal referents in subject position in developmental perspective.

Zero subjects in acquisition

Grammatical subjects constitute a heterogeneous category, consisting of different thematic roles that are related to transitivity values, word class, or even quirky, non-nominative case marking (Barðdal, 2004; R. M. Dixon, 1994; Dowty, 1991; Lambrecht, 2000; Li & Thompson, 1976; Mohanan & Mohanan, 1990). As such, its development should be analyzed relative to its various forms and functions in a specific language (Dattner et al., 2019). And interwoven with the array of forms and functions designated as "grammatical subjects" is the possibility, in some languages, of a zero or null subject. Two major perspectives on null subjects are (i) subjectless constructions and (ii) zero realization of pronominal subjects.

The first perspective involves studies on subjectless constructions that have no formal slot for a grammatical subject and typically lack finite predicates, which occur in typologically different languages such as Korean or Hebrew (Halevy, 2020; Park, 2001). They are mainly used to express non-agentive, non-definite, impersonal, often modal functions (Berman, 1980, 2011; Blevins, 2013). See the following Polish example, cited from Kibort (2011):

(1) *Było* *codziennie* *sprzłątane.*
was.3rd.Sg.N every-day tidy.Part.Sg.N
'[It] has been cleaned every day. / There has been cleaning every day.'

This type of null subject clauses with no subject slot is excluded from the current analysis. Rather, we target the second perspective on zero subject phenomena – namely, the seemingly optional zero realization of pronominal subjects in finite contexts as in the following examples from Hebrew (2) and Spanish (3):

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

(2) *ata/ ∅* *pagáta* *be-shalosh* *gulot* *ktanot.*
you/ ∅ hit.2nd.Sg.masc.past at-three marbles little.
'You hit three little marbles.'

(3) *Ella/ ∅* *llegó* *de* *Madrid.*
She/ ∅ arrived.3rd.Sg from Madrid.
'She arrived from Madrid.'

Zero realization of subjects is not possible in every language, leading to the vastly studied topic of the null subject parameter. Within the generative search for a universal grammar (UG), zero subjects are mostly regarded as an omission phenomenon (Hyams & Wexler, 1993; but see also Uziel-Karl & Berman, 2000 for a non-generative omission-based account). Given that only some languages (termed Pro-Drop languages) allow a zero realization of the pronominal subject, two questions are usually asked in the literature: first, how and when do children set their subject-related parameter within UG; and second, why do children acquiring a non pro-drop language leave out sentential subjects in their early development (Valian, 1991). The search for answers to the early omission question has mainly revolved around children acquiring English (as a canonical non-pro-drop language) versus Spanish or Italian (as canonical pro-drop languages). Within the generative tradition, explanations of this developmental phenomenon ranged from a mis-setting of the parameter as pro-drop (Hyams, 1986), through biological maturation (Borer & Wexler, 1987), to performance factors (Valian, 1991). More recent generativist accounts of the acquisition of grammatical subjects have used minimalist approaches to explain the null-subject parameter (e.g., the various works in Biberauer, Holmberg, Roberts, & Sheehan, 2009). Revisiting her original and modified proposals, Hyams (2011) suggests a less modular account where general principles of information structure, constrained by the grammatical system, explain the acquisition of pronominal subjects.

Outside the generative tradition, the acquisition of zero and pronominal subjects has been treated as an alternation (rather than omission) phenomenon. Children growing up in languages that allow such alternations need to pay attention to the grammatical, semantic and discursive contexts in which predicates and their subject arguments are marked in relation to each other. With focus on Spanish and Italian, accounts of learning zero/pronoun alternations range from working memory to reference tracking in discourse, including sociolinguistic variables (Shin & Cairns, 2009; Zahler, 2018). Orozco (2016) showed that the alternation between pronominal and zero subjects in Xalapa Mexican Spanish is related to the subject's person, number and gender, as well as to reference chains, verb semantic class and temporality. These developmental studies emphasize the role of input and frequent usage patterns in the ambient language. For example, Paradis and Navarro (2003) show that monolingual Spanish-speaking children appropriately use zero and pronominal subjects at two years. This conclusion is shared by Orvig et al.'s (2010) study of clitic pronouns in dyadic conversations, suggesting that children's uses of pronouns reflect early pragmatic skills. Similarly, Serratrice (2005) shows that Italian-speaking children can use null and explicit subjects in pragmatically appropriate ways early on, with more explicit pronominal subjects increasing with MLU, mainly in first person. Shin (2016) argues for a piecemeal learning of zero and pronominal subjects in Mexican Spanish, where frequency and pattern simplicity drive acquisition. And Forsythe, Greeson, and Schmitt (2022) show that subject realization alternations are probabilistically associated with appropriate reference in a particular order: children first

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

associate null subjects with continued reference to the preceding subject antecedent, and overt subjects with a switch in reference. Only after children have tracked the conditions governing subject pronoun alternations in their input and master them for a subset of the subject category do they go on to generalize it to the entire category (Forsythe et al., 2022).

These findings are of relevance for the present study on Hebrew, investigating how children develop zero vs. pronominal subject expression in their peer talk. Hebrew is considered a "mixed" pro-drop language with regards to subject pronoun expression (Berman, 1990; Shlonsky, 2009; Uziel-Karl & Berman, 2000), as it allows zero reference to subjects only in part of the paradigm (see below). Thus, children need to learn not only the morpho-syntactic conditions where the zero and pronominal subjects are possible, but also the communicative contexts under which they take place.

Hebrew pronominal subjects, first and second person

The current paper explores patterns of pronominal/zero subject expression in clauses with past and future tense verbs marked for 1st and 2nd person agreement, a research focus that is based on general discourse, as well as Hebrew-specific motivations. From a general perspective, all languages have 1st and 2nd person pronouns (Dixon, 2010) referring to the deictic speaker and addressee roles in conversation, which are always animate and human. These are inherently different from 3rd person pronouns, referring to less accessible entities, including non-animate and even abstract reference (Ariel, 1999, 2001; Charnavel, 2019). From a developmental perspective, Forsythe et al. (2022) accordingly show that children acquire the referential distribution of 1st and 2nd person pronominal subjects before generalizing to the 3rd person. Indeed, the two deictic roles of 1st and 2nd person frequently occurred in our data in the children's peer talk and had both prominence and accessibility in their natural, spontaneous conversations.

From a Hebrew-specific perspective, the only morpho-syntactic contexts that fully allow pronominal and zero subjects, with no grammatical or pragmatic restrictions, are 1st and 2nd person in past and future tense, as detailed in Table 1. Therefore 3rd person pronominal subjects were excluded from the current analysis. Support for this focus can be found by looking at adult Hebrew usage of the grammatical subject. The distribution of zero and pronominal subjects in adult Hebrew is reported in Ariel (1999), who shows that 3rd person accessible referents are highly infrequent, meaning that the potential for 3rd person zero realization as a result of pragmatic consideration is very low. Two data types provide the following distributions: in long phone conversations, 51.3% of zero and pronominal subjects are 1st person, 35.7% are 2nd person, and only 13% are 3rd person. Within these conversations, 31.55% of zero subjects are 1st person, 60.7% are 2nd person, and only 7.7% are 3rd person. In face to face conversations, zero and pronominal subjects consist of 41.2% 1st person and 41.2% 2nd person, while 3rd person accounts for only

Table 1. Non-restricted zero realization of Hebrew grammatical subjects by verb tense and subject person in Hebrew

	1st	2nd	3rd
Past	✓	✓	✗
Present	✗	✗	✗
Future	✓	✓	✗

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

17.5%. These findings motivate the exclusion of 3rd person pronouns from the current study: while 3rd person zero realization of subject is possible, it is highly limited, given specific conditions that involve complex syntactic, morphological and discursive factors (Gutman, 2004; Melnik, 2007).

Our range of data is also temporally restricted, as mentioned above, to past and future tense. Present-tense verbs were excluded from the current analysis, as markedly differing from past and future tense verbs in generally requiring pronominal subjects, while not allowing zero subjects. This is because past and future Hebrew verbs agree with their subjects in number, gender and person, while present tense verbs agree only in number and gender (excluding person). Moreover, with regards to future tense, we included only those verb tokens that could be clearly identified as carrying future semantics rather than the imperative sense, as detailed in the method section below.

Hebrew zero and pronominal subjects in acquisition

Many studies on Hebrew pronominal acquisition were carried out within the generative framework, concerning questions of pro-drop and root infinitives (uninflected verbs, see Grinstead, 2016; Lustigman, 2015), or subjects in relative clauses (e.g., Friedmann & Novogrodsky, 2004). As Hebrew is considered a mixed pro-drop language, it poses a problem for a unified syntactic explanation (Vainikka & Levy, 1999). Much of this research was devoted to an intra-theoretic debate regarding the status of different syntactic categories and processes (e.g., Armon-Lotem, 2008; Borer, 1989; Friedmann & Novogrodsky, 2004; Hacohen & Schaeffer, 2007; Levy & Vainikka, 2000; Schaeffer & Ben Shalom, 2008; Vainikka & Levy, 1999). Levy and Vainikka (2000) showed that children's omission patterns approximate adult language before age 2, with MLU around 2.5. Vainikka and Levy (1999) proposed a syntactic principle to explain the Hebrew (and Finnish) developmental data, such that the syntactic position of subject-verb agreement features varies cross-linguistically and within language. They argued that the specific location of the agreement feature, together with a pragmatic idea of recoverability, constitute a unified, cross-linguistic explanation.

In a series of studies aimed at pinpointing the structural location of subject-related phenomena, Schaeffer and colleagues examined the choice of nominal expressions for grammatical subjects. Hacohen and Schaeffer (2007) examined pragmatically inappropriate Hebrew subjects in bilingual Hebrew–English development. Arguing that Hebrew verbs allow null subjects, they showed that bilingual and monolingual children produced similar subject-verb agreement patterns. According to Schaeffer and Ben Shalom (2008), grammatical subjects are deictically anchored, requiring a connection between the syntactic feature of person and discourse participation, together with assumptions about the mental states of speaker and hearer. They propose that clear, productive person marking in the adult language constitute the trigger for this early acquisition of pragmatically appropriate choices of nominal expressions.

Armon-Lotem's 2008 account of the interrelationship between the acquisition of verbal morphology and the use of explicit and null subjects in Hebrew is of importance to the current analysis. Her study underscores the relative order between the early use of verbal morphology and of different types of grammatical subjects, taking it as evidence for the central role of the finite verb in acquiring the morpho-syntactic features of the subject.

From a different perspective, Berman's 1990 seminal analysis of Hebrew subject acquisition emphasized children's ability to rely on different types of evidence in

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

morpho-syntactic acquisition. Among the different null subject phenomena analyzed by Berman, the relevant type for the present paper is what she terms "agreement marking in simple clauses." Berman notes that while null subjects are grammatically licensed with verbs that have rich agreement marking (person, number, and gender), the acquisitional picture is more complex, as children need to pay attention not only to the structural cues in the inflectional paradigm, but also to situational licensing cues.

In a study of young children's use of subject (and object) ellipsis in Hebrew (ages 1;5–2;4), Uziel-Karl and Berman (2000) show that subject ellipsis is motivated mainly by pragmatic factors at first, while development is manifested by adhering to the morpho-syntactic rules of Hebrew grammar. Taking a phase-based perspective to describe development, Uziel-Karl and Berman term the initial pragmatically governed period as involving elementary criteria. These include conditions that are necessary for preventing communication breakdown. Conversely, the later, grammar-based period involves advanced criteria, including adult-like morphosyntactic knowledge that provides sufficient conditions to prevent ungrammaticality. Finally, and with relevance to the present paper, children integrate the two types of knowledge at the most mature phase, using ellipsis to meet appropriate discourse functions. This period is covered in the present analysis of children at the ages of 2;0–8;0. Uziel-Karl and Berman (2000) emphasize that they "do not side with either strictly grammatical or discourse-based accounts but suggest instead that both sets of factors play a role" (p. 468). In the present paper we concur with this approach, showing that these are not two opposite governing criteria, but rather complementary usage conditions for the manifestation of zero and pronominal subject in Hebrew child language.

In sum, three significant findings arise from the literature review on zero and pronominal subject acquisition in Hebrew. First, 1st and 2nd person subjects are inherently different from 3rd person subjects, both in their grammatical paradigm and in their usage patterns. Second, the acquisition of verb temporality and agreement inflection is crucial for the development of pronominal subjects. Third, pragmatic contexts affect grammatical subject realization. In light of these findings, we formulate the current paper's aims and hypotheses.

Aims and hypotheses

The research framework

From a usage-based, developmental psycholinguistic perspective, four different assumptions account for the path whereby grammatical subjects and their realizations emerge and evolve in children. First, relating to the theoretical status of the zero subject. Most previous studies have focused on the circumstances allowing the omission of grammatical subjects (Vainikka & Levy, 1999), where "speakers can choose to either express a subject pronoun or omit it" (Shin, 2016). Omission is a theory-dependent notion, presupposing a complete syntactic structure in the speaker's mind that needs to be given overt expression, from which a sparser or less well formed structure is derived. But in a usage-based theoretical perspective, both zero and pronominal subjects are construed as complete discursive-dependent communicative acts (Ariel, 2001; Du Bois, 1987, 2003). Different structures (e.g., word order or referent realization) are associated with different usage conditions and discursive functions, creating discursive usage patterns. In this way, a zero realization of a referent does not involve an omitted pronoun, but is rather a different structure used in specific discourse contexts.

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

A second assumption is that acquisition of grammatical subjects should be explored within the usage patterns of a specific language typology, taking into account the distributions of explicit and consistent as well as less equivocal and less consistent phenomena in the language (Berman, 1980; Conwell, 2020; Dattner et al., 2019; Serratrice et al., 2004). These distributional properties constitute cues that children pick up in the ambient language and relate to semantic information in the construal of grammatical subjects.

Thirdly, all previous studies investigating early syntax in morphologically rich languages such as Hebrew emphasize the co-development of verb tense and agreement with the grammatical subject (Armon-Lotem, 2008; Avram, 2018; Grinstead, 2000; Hyams, 2011; Veneziano & Clark, 2016). Therefore, these two evolving phenomena should be studied together in children's productions.

A final and critical assumption in the current study is that the discursive contexts that enable children to produce grammatical constructions should be part of this investigation (Alonso-Ovalle et al., 2002; Serratrice, 2005). These links between grammar and discourse are conceptualized as Discourse Profile Constructions in Dattner (2015, 2019) for adult language, and in Dattner et al. (2019) in a developmental perspective. For example, Dattner (2019) shows that the usage patterns of the dative case in spoken adult Hebrew are best explained as specific links between grammar and discourse. Consider the example in (4), where the two dative occurrences are the syntactic arguments of the same verb type (*la-tet* 'to give') – however, they behave differently since they belong to two different usage patterns, each linked to a different discursive function and dative meaning:

(4)	*tnu*	*lánu*	*beynatayim*	*latet*	*lahem*	*darga zmanit.*
	give.2nd.pl.Imperative	us.Dative	meanwhile	give.Inf	them.Dative	position temporary.

'For now, let us give them a temporary position.'

The first dative (*tnu lanu* lit. 'give us = let us') is interpreted as an enabled/allowed participant linked to a usage pattern consisting of transitive verbs with non-accusative direct objects. The second dative (*latet lahem* 'to-give them') is interpreted as a recipient linked to a different pattern of transitive verbs with accusative direct objects. These examples demonstrate the need for the *Discourse Profile Construction Hypothesis*: meaning is usage-based (Wittgenstein, 1953), and a particular construal of a state of affairs is conveyed through a specific usage pattern.

Usage patterns and Discourse Profile Constructions in acquisition

Dattner (2015, 2019) and Dattner et al. (2019) show that the frequent usage of a particular grammatical construction within its discursive context may yield a coherent, multifactorial usage pattern, defined as a Discourse Profile Construction. This idea extends the Argument Structure Construction notion (Goldberg, 1995) and the Behavioral Profiles notion (Divjak & Gries, 2006; Gries, 2010) to include discursive and communicative functions, based on statistically defined similarity between tokens of language use. That is, Discourse Profile Constructions are assumed to be basic discourse-level constructions, an extension of Argument Structure Constructions which are basic syntactic-level constructions (Goldberg, 1995).

As illustrated in (4), Dattner (2015, 2019) and Dattner et al. (2019) demonstrate that in terms of language use, morphological and syntactic usage patterns may correspond to

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

specific discursive and communicative functions, beyond their semantic/pragmatic functions. The Discourse Profile Construction Hypothesis argues that communicative functions are correlated with sets of parameters conveying morphological, syntactic, pragmatic, and conversational information. These correspondences emerge from the conventional nature of usage, such that the same function tends to be served by similar utterances in terms of co-occurring linguistic categories. A usage pattern is thus a set of linguistically relevant categories that are shared across a set of utterances, constituting high similarity between these utterances.

Applying this notion to child language acquisition, we are interested in finding whether the use of pronominal and zero subjects in Hebrew is related to the properties of the subject referent (number and person), to verb tense (representing temporality), and to the role of the utterance in the flow of conversation. If these features indeed form patterns of usage, we look for the discursive or communicative functions served by these patterns. Thus, we search for statistically defined sets of tokens in the corpus that are coded in a similar manner, yielding identifiable clusters of similar tokens and of linguistic properties that tend to be used together.

A specific motivation for this course of action comes from Dattner et al. (2019) who showed that children's developing use of the range of Hebrew lexical and pronominal subjects in clauses with different word orders can be explained using the Discourse Profile Construction hypothesis. For example, one usage pattern found in Dattner et al. (2019) concerns the grammatical features of post verbal lexical subjects with present tense verbs. This pattern was shown to be probabilistically linked to a specific discursive function: introducing new topics to the discourse and referring to less accessible entities outside of the conversational arena. This Discourse Profile Construction (demonstrated in (5)) is used appropriately in conversation only from age 5;0 and onward.

(5) *li,* *le-Ana'el* *ve-le-Ofri* *yesh* *manuy* *le-hacagot.*
to.me, to-Ana'el and-to-Ofri be subscription to-shows.
'I, Ana'el, and Ofri have a theater subscription.'

Against this background, the present paper investigates the development of Hebrew 1st and 2nd person grammatical subjects with their zero or pronominal realization as part of the grammatical inventory the Hebrew-speaking child needs to acquire. Given the ages of our peer talk participants (2;0–8;0 years), our corpus does not include pre-grammatical subjectless root infinitives (Lustigman, 2015). We thus analyze only utterances with clearly marked finite verbs, accounting for the realization of their relevant grammatical subjects.

Research objectives

Forsythe et al. (2022) define the learning problem regarding null vs. overt pronominal subjects as acquiring the "probabilistic association between pronoun realization and pronoun reference." The present paper approaches this matter from a different perspective, asking what are the usage patterns of zero and pronominal subjects in development, not as alternations but rather as belonging to different Discourse Profile Constructions, as defined above. Arguing for a Usage-Based linguistic competence, Tomasello (2003) defines it as "the mastery of a structured inventory of meaningful linguistic constructions" (p. 99). Accordingly, we define the learning problem as mastering the set of form-function

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

correlations (at various levels) that involve zero and pronominal subjects. For example, we search for those concrete patterns of language use in which a zero subject is used with reference to particular discourse entity, accompanied by a verb with a particular tense, and show that it has a particular function, while a different pattern (e.g., zero subject referring to another entity and accompanied by a verb inflected to a different tense) has a different function. That is, we argue that it is not a matter of learning the statistical distribution of "the subject" relative to its referents, but rather mastering different constructions for which a subject is relevant.

The present paper sets out two objectives, corresponding to two sets of analyses and results. First, we look for the age-related data distribution of zero and pronominal subjects by verb tense, number and person, as well as by conversational utterance type. To the best of our knowledge, this rigorous, fine-grained analysis has not been carried out to date in studies concerning the development of Hebrew grammatical subjects. A second set of analyses aims to show that the development of the usage of zero and pronominal subjects in Hebrew can be best explained using the Discourse Profile Construction hypothesis.

We expect to find usage patterns which are sets of utterances, clustered together according to statistically based similarity. We hypothesize that the distribution of usage patterns in the corpus will not be uniform across age groups: that is, for each age range there should be a pattern that is used significantly more frequently than others, serving a different discursive-communicative function. Development is hypothesized to be evident in the usage variation of pronominal and zero subjects, such that with age, each construct will be used in more environments.

Having identified usage patterns in the data, we follow up with analyses seeking the correspondence of usage patterns with unique discursive functions, thus constituting Discourse Profile Constructions for whose usage we aim to find developmental paths. Crucially, we do not presuppose a set of discursive functions that may correspond with the resulting usage patterns. Rather, we work from the bottom-up: given a usage pattern of co-occurring linguistic parameters and a set of similar language tokens, we aim to interpret their shared discursive function.

Data and Method

Data

The data for the present paper is the peer-talk corpus (Zwiling, 2009), composed of the recordings of 54 children (equal number of boys and girls) in six age groups: 2;0–2;6, 2;6–3;0, 3;0–4;0, 4;0–5;0, 5;0-6;0, 7;0–8;0. The children were engaged in triadic conversations, three triads per age group, totaling eight hours of recordings. All participants were native Hebrew speakers from medium-high SES, with no language disorders or other developmental problems. Sessions were 30 minutes long, except for those of the youngest groups (ages 2;0–2;6), which were 40 minutes long each, culminating in 1.5 hours of recordings for each age group (30 minutes $\times$ 3 triads), and two hours for the youngest group (40 minutes $\times$ 3 triads). Sessions were conducted in the children's homes, audio-recorded and transcribed. The recordings yielded 36,490 words in 11,870 utterances altogether. Note that it is not the case that every utterance in the corpus had a subject-relevant environment that could be included in the data. For example, clauses with verbs in the imperative of infinitive inflections are irrelevant for subject coding in Hebrew. Table 2 provides the distribution of subject types in the general corpus, consisting of 1st,

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

Table 2. Subject type distribution in the corpus: Lexical, pronominal and zero subjects in future, present, and past tense, singular and plural

	1st person		2nd person		3rd person		
	Null	Pronominal	Null	Pronominal	Lexical	Pronominal	Total
Future							
Plural	263	17	8	7	20	2	317
Singular		257	20	28	83	80	468
Past							
Plural	74	15	11	3	36	30	169
Singular	368	133	81	35	154	203	974
Present							
Plural		136		52	246	36	470
Singular		948		406	902	1041	3297
Total	705	1506	120	531	1441	1392	5695

2nd and 3rd person, future, past, and present tenses, singular and plural, with lexical, pronominal and null realization. These accumulate to a corpus of 5,695 tokens.

The current paper examines a sub-corpus composed of all utterances that are prone to zero and pronominal realization, with no clear restrictions. As shown earlier, these are non-lexical subjects, in first and second person, with past and future tense verbs (gray colored cells in Table 2). The final corpus analyzed for the purposes of the current study thus consisted of 1,320 utterances, distributed across the age groups as detailed in Table 3.

Method

Coding variables

We coded for the following variables: (i) age group: 6 age groups: 2;0-2;6, 2;6-3;0, 3;0-4;0, 4;0-5;0, 5;0-6;0, 7;0-8;0, (ii) grammatical subject realization: pronoun vs. zero (i.e., verb agreement marking only), (iii) grammatical subject number: singular vs. plural, (iv) grammatical subject person: first vs. second, and (v) verb tense: past vs. future.[1] This coding scheme is exemplified in the following (examples are drawn from the current corpus unless stated otherwise):

(6) *nafálti* 'I fell' (2;0-2;6, zero pronoun, Sg, 1st, Past tense)

(7) *ani lakáxti otam* 'I took them' (2;0-2;6, full pronoun, Sg, 1st, Past tense)

[1]Future tense in 2nd person and imperatives in Hebrew can sometimes take exactly the same form (e.g., *tedaber* 'talk!/you.will talk'). Since zero subject is mandatory in imperatives, the coding scheme made sure that imperatives were not included in the current analysis. Thus, we excluded all future-formed imperatives, based on a pragmatic analysis (involving the two authors) of each token. Only those cases that were finally counted as Future rather than imperative are included in the present paper.

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

Table 3. Corpus size: Number of zero/pronominal subjects per age group, for the same recording times

Age group	2;0–2;6	2;6–3;0	3;0–4;0	4;0–5;0	5;0–6;06	7;0–8;0	Total
Tokens	105	124	209	230	307	345	1320

This set of coding variables includes the relevant features for pronominal subjects, based on the literature discussed above and the morphosyntactic features of Hebrew.

Serratrice (2005) shows that children's conversational skills develop in tandem with grammatical acquisition, and Dattner et al. (2019) showed that the use and development of the grammatical subject is tightly linked to the role of the utterance in conversation, especially in peer interaction. Children engaging in peer talk are active participants in communicative events, and must consider the role they play in conversation and discourse (Forrester & Cherrington, 2009). Thus, in addition to grammatical features, and following Dattner et al. (2019), each utterance was coded for the role it played in the children's conversation. Note that this is not a pragmatic-based coding, but rather a conversational coding aiming at highlighting the changing use of subjects within the conversational streamline. We coded the following four conversational roles:[2]

- Initiating utterances: utterances which initiate a conversational sequence, to the extent that these type of utterances shows no interaction with the previous utterance. The utterance immediately preceding an Initiating utterance is irrelevant in terms of structure, content, topic, or subject referent, as shown by B's utterance in the following sequence:

(8) A: *koev li.* 'I'm hurt.'
B: *ani yavi et hakarit hagdola.*'I'll bring the big pillow.'
(B initiates an utterance, irrelevant to what A said)

- Repeating utterances: utterances which fully repeat (i.e., verbatim) the utterance immediately preceding them, as shown by B's utterance in the following sequence:[3]

(9) A: *bo neshev kan.* 'Let's sit here.'[4]
B: *bo neshev kan.* 'Let's sit here.'
(B repeats A's utterance)

- Resonating utterances: utterances that resonate the structure of the utterance immediately preceding them. Resonating utterances show some interaction with

[2]The present coding scheme captures the relevant pragmatic features of switch topic and focus, while other features (such as salient referent, person, or animacy) were irrelevant due to our focus on 1st and 2nd person, which are marked for person, and are always human and salient.

[3]Given the usage-based framework of the current paper, full repetitions were included in the analysis since these are language events that constitute genuine phenomenon in children's conversations and form part of the children's experience with language.

[4]Note that some sentences might have two predicates; for example, *let* and *sit* in the present case (see also Ex. (10)). Within the framework of our coding scheme, sentences with two predicates (complex or not) were analyzed as two separate tokens regarding subject realization. Thus, in the present case, the zero subject realization of the imperative *let* was excluded from the current data, while the zero realization of the first person plural future verb *sit* was included.

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

previous utterances, moving on with the flow of conversation by re-using a recent structure, and have a critical role in building coherence in extended dialogue (del Prado Martín & Du Bois, 2015; Du Bois, 2014). Consider B's utterance in the following sequence:

(10) A: *bo na'ase et ze.* 'Let's do [=make] it.'
B: *bo na'ase makel ksamim.* 'Let's make a magic stick.'
(B is using the same structure as A, with a slight variation)

- Response utterances: utterances which respond to the immediately preceding utterance. These utterances show full interaction, and fully use the to and fro nature of conversation, as B does in the following sequence:

(11) A: *beyntaym ani yitle et ha-balonim.* 'In the meantime, I'll hang the balloons.'
B: *ve-ani yaxin gam uga.* 'And I will bake a cake too.'
(B is responding to A's utterance)

Revealing usage patterns through cluster analysis

To account for the usage patterns of zero and pronominal subjects we applied a Hierarchical Clustering on Principal Components (Husson, Josse, & Pagès, 2010; Lê, Josse, & Husson, 2008; R Core Team, 2019), along the lines presented in Dattner (2019) and Dattner et al. (2019). This procedure was first used to reveal non-age related multifactorial usage patterns in the corpus, similar to Divjak and Gries's (2006) behavioral profiles. The cluster analysis yielded distinguishable usage patterns in the corpus, based on correspondences between the coding variables. Following the cluster analysis, we performed a Correspondence Analysis between the resulting clusters and the age groups in our data to see how the usage patterns were distributed across the ages.

Finally, we looked for a common function served by each usage pattern. This was done to define Discourse Profile Constructions so as to explain the development of zero and pronominal Hebrew subjects in terms of the discursive and communicative functions linked to their usage patterns.

Results

Distribution of variables across age groups

Figure 1 presents the distribution of the study variables across age groups. A chi-square test of independence was performed to examine the relation between age and each of the variables. The relation between age and grammatical subject type (top-left panel in Figure 1) was significant ($x^2(5, N = 1323) = 21.875, p = .0001$): A Pearson residual analysis (Delucchi, 1993) reveals that the 2;0-2;6 age group used more zero subjects and fewer pronominal subjects than would be expected by chance. The relation between age and number (top-right panel) was also significant ($x^2(5, N = 1323) = 15.328, p = .009$): A Pearson residuals analysis shows that the 2;0-2;6 age group used significantly more plural subjects than would be expected by chance. The relation between age and tense (mid-left panel) was significant ($x^2(5, N = 1323) = 28.507, p < .0001$): A Pearson residuals analysis shows that the 3;0-4;0 age group used fewer past tense verbs and more future tense verbs than would be expected by chance. The relation between age and person (mid-right panel)

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

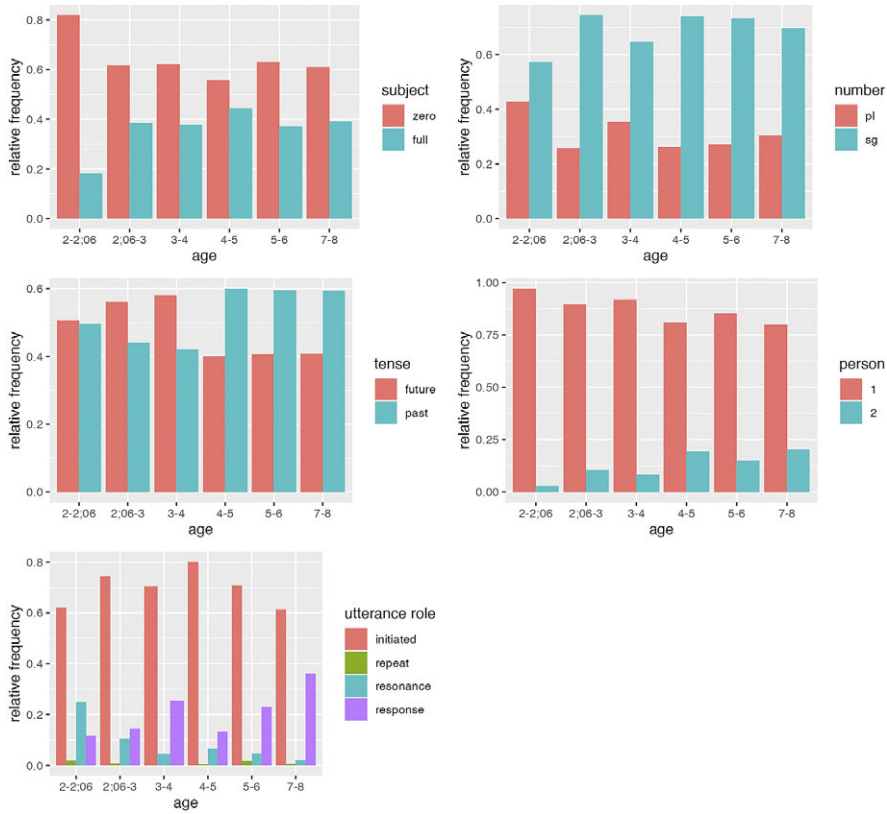

Figure 1. Relative frequencies of variables across age groups

was significant as well ($x^2(5, N = 1323) = 33.025, p < .0001$): A Pearson residuals analysis shows that the 2;0-2;6 and 3;0-4;0 age groups used fewer 2nd person subjects than would be expected by chance, and that the 7;0-8;0 group used more 2nd person subjects than would be expected by chance. Finally, the relation between age and utterance type (bottom panel) was significant ($x^2(15, N = 1323) = 132.794,\ p < .0001$): A Pearson residuals analysis shows that (i) the 2;0-2;6 age group used more resonating utterances and fewer response utterances than would be expected by chance, (ii) the 2;6-3;0 and 4;0-5;0 age groups used fewer response utterances than would be expected by chance, and (iii) the 7;0-8;0 age group used more response utterances and fewer resonance utterances than would be expected by chance.

Subject realization across variables and age groups

Having presented the general distributions of the variables, we now examine the realization of grammatical subjects as zero vs. pronominal relative to the other variables in the data, as presented in Table 4.

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

Table 4. Subject realization: Tense, person, and number (proportions)

tense	person	number	subject	2;0-2;06	2;06-3;0	3;0-4;0	4;0-5;0	5;0-6;0	7;0-8;0
future	1	pl	zero	0.98	1	0.93	0.93	0.96	0.87
			pro	0.02	0	0.07	0.07	0.04	0.13
		sg	zero	0	0	0	0	0	0
			pro	1	1	1	1	1	1
	2	pl	zero	0	0	0	0	0.36	1
			pro	0	0	0	0	0.64	0
		sg	zero	1	0.5	0.6	0.22	0.71	0.28
			pro	0	0.5	0.4	0.78	0.29	0.72
past	1	pl	zero	0.67	1	1	0.75	0.85	0.8
			pro	0.33	0	0	0.25	0.15	0.2
		sg	zero	0.87	0.9	0.68	0.62	0.71	0.78
			pro	0.13	0.1	0.32	0.38	0.29	0.22
	2	pl	zero	0	1	1	0.67	1	0.75
			pro	0	0	0	0.33	0	0.25
		sg	zero	0.5	1	0.91	0.53	0.81	0.68
			pro	0.5	0	0.09	0.47	0.19	0.32

Table 4 shows that a large number of cells have a zero-frequency, indicating that many of the variables are strongly correlated. That is, some variables have a very restricted pattern of usage in the corpus. For example, 2nd person plural subjects in future tense are almost never realized as a full pronoun, except in the 5 year olds. In fact, future tense 2nd person plurals are a very late bloomer phenomenon, as both their pronominal and zero realizations are found only at the two oldest groups, as in the following examples:

(12) *atem* *te'argenu* *et* *ha-uga.*
You.pl will.organize ACC the-cake.
'You will set the cake.'
(pronominal subject, 5;0-6;0)

(13) *she-ted'u!*
That-you.pl.will.know!
'So you'll know!'
(zero subject, 7;0-8;0)

Another interesting correlation is the 2nd person plural subjects in past tense, which are almost always realized as zero, as in the following:

(14) *amártem* *be-oto* *zman.*
say.2nd.pl.masc.past at-the.same time.
'You said it at the same time.'
(zero subject, 7;0-8;0)

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

Another finding apparent in Table 4 is the descending number of zero-frequency cells with age. That is, the variety of possible combinations of subject features rises with age. This indicates more varied grammatical environments in older vs. younger children. In sum, Table 4 shows that many of the variables correspond with one another.

Hierarchical Clustering on Principal Components: Usage patterns of zero and pronominal subjects

The findings presented above suggest that a correspondence-based analysis is apt for our study aims. Thus, we now move on to analyzing the data through the perspective of similarity between utterances, given the correspondences between the coding variables. Crucially, the age group variable was not included in the computation of similarity, allowing us to account for correspondences between usage patterns and age groups. The cluster analysis yielded five clusters, three of which account for most of the data, as shown in Table 5.

The description of variables in each cluster is given in Tables 6–7, as follows: The first column lists the relevant coding categories. The second column (Belong To Cluster) presents the relative frequency of tokens in the corpus that are coded for the relevant category, which belong to the cluster. For example, Table 6 shows that 100% of Resonance utterances in the corpus belong to cluster 2. The third column (Within Cluster) shows the relative frequency of the tokens in the cluster which are coded for a specific category. For example, within cluster 2, 100% of the tokens are Resonance utterances. The fourth column (Global) shows the relative frequency of each category in the corpus, regardless of cluster. The two last columns, *p*.value and *v*.test, indicate the goodness of representation of each category within the cluster. A positive *v* .test indicates that the category is

Table 5. Cluster sizes

Cluster	Tokens	Proportion
1	341	0.26
2	84	0.06
3	11	0.01
4	387	0.29
5	500	0.38

Table 6. Description of variables in clusters 2, 3

Category	Belong To Cluster	Within Cluster	Global	*p*.value	*v*.test
Cluster 2					
utterance type = Resonance	100.00	100.00	6.35	0.00	24.75
age = 2;0-2;6	24.76	30.95	7.94	0.00	6.44
Cluster 3					
utterance type = Repeat	100.00	100.00	0.83	0.00	10.85

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

Table 7. Description of variables in clusters 1, 4, 5

Category	Belong To Cluster	Within Cluster	Global	*p*.value	*v*.test
Cluster 1					
number = pl	85.46	100.00	30.16	0.00	34.21
subject = zero	38.14	92.38	62.43	0.00	14.37
tense = future	43.36	76.54	45.50	0.00	13.52
person = 1	29.47	97.65	85.41	0.00	8.48
utterance type = Initiated	30.47	82.11	69.46	0.00	6.07
Cluster 4					
tense = past	51.73	96.38	54.50	0.00	21.56
subject = zero	43.95	93.80	62.43	0.00	16.50
number = sg	40.37	96.38	69.84	0.00	15.15
utterance type = Initiated	40.04	95.09	69.46	0.00	14.39
person = 2	65.28	32.56	14.59	0.00	11.31
Cluster 5					
subject = pronoun	83.70	83.20	37.57	0.00	27.69
number = sg	52.92	97.80	69.84	0.00	19.32
utterance type = Response	74.11	45.80	23.36	0.00	14.93
tense = future	46.18	55.60	45.50	0.00	5.74
person = 1	40.00	90.40	85.41	0.00	4.09

over-represented in the cluster. Tables 6–7 report only those variables with a positive *v*. test and significant *p*.values.

The two very small and restricted clusters 2 and 3 respectively comprise all Resonance utterances (cluster 2) and Repeat utterances (cluster 3) in the data (Table 6). That is, the usage patterns associated with Repeating and Resonating utterances are significantly different from those of the more cooperative and conversational Initiated and Response utterances.

The cluster analysis yielded three main clusters as usage patterns consisting of grammatical form and conversational role, reported in Table 7.

Cluster 1 comprises 26% of the data. It is a cluster of Initiated utterances in future tense, referring to 1st person plural entities, realized as zero subjects. An example of this cluster is:

(15) *bo* *neshev* *kan.*
come sit.1st.pl.fut here.
'Let's sit here.'

Cluster 4 (Table 7) comprises 29% of the data, depicting Initiated utterances in the past tense, referring to 2nd person singular entities, realized as zero subjects, as in the following:

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

(16) *pagáta* *be-shalosh* *gulot* *ktanot.*
hit.2nd.sg.masc.past in-three marbles little.
'You hit three little marbles.'

Cluster 5 is the largest in the corpus (38%). It concerns Response utterances with pronominal subjects, referring to 1st person singular entities (i.e., the speaker), in future tense. An example of this cluster is:

(17) *ani* *ya'azor* *lexa,* *tov?*
I help.1st.sg.fut to.you, OK?
'I will help you, OK?'

The next step was to look for correspondences between the three usage patterns and the age groups in our data.

The developmental path: Usage patterns and age

We calculated the correspondence between the cluster each utterance belonged to, and the age group that produced it. Figure 2 presents the results of a Correspondence Analysis (CA).

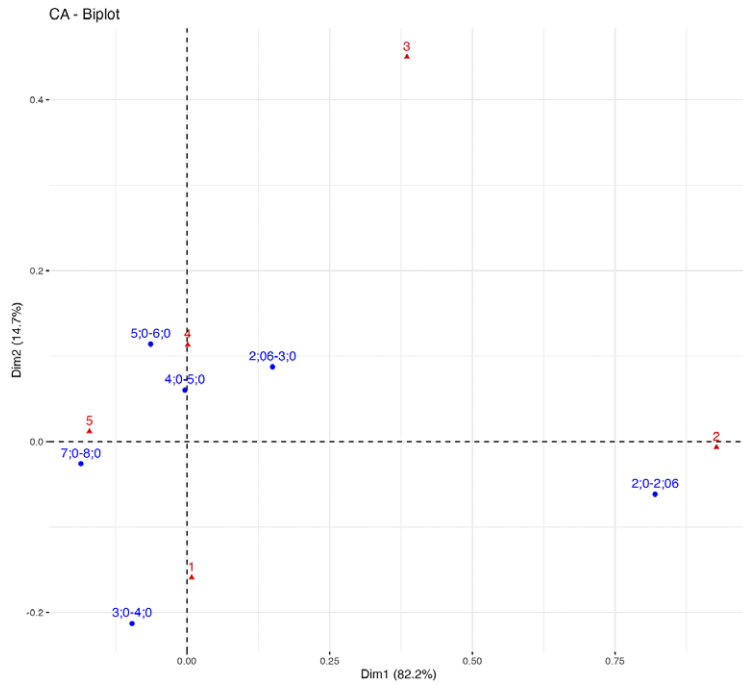

Figure 2. Correspondence Analysis of cluster and age. Clusters are represented as red colored numbers; age groups appear in blue. Two age groups are located closer to each other if they use the same cluster, and two clusters are located closer to each other if they are used by the same age group.

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

Figure 2 shows that the usage patterns in the data, represented as clusters of utterances and linguistic parameters, are not evenly distributed across the age groups. Rather, a clear developmental path emerges, as follows: The youngest age group in our data (2;0-2;6) is located at a great distance from the other groups, indicating the small number of utterances that share usage patterns with older ages. The next closest group is the consecutive 2;6-3;0, located between the youngest group and the rest of the age groups as a bridge between very young usage with regard to zero and pronominal subjects and more mature usage patterns. The 3;0-4;0 age group is located at the bottom left quarter, slightly isolated relative to the adjacent locations of the older groups. This can be interpreted as yet another bridge between young and mature patterns, this time with an inclination towards the more mature groups: the group's relative adjacent location to the older rather than to the younger groups indicates that they share more usage patterns with their older peers than with their younger ones. The 4;0-6;0 groups are very close to each other, sharing a very similar use of zero and pronominal subjects, and different from their younger 3;0-4;0 group. Finally, the oldest group (7;0-8;0) is located in between the 3;0-4;0 and the 4;0-6;0 groups, indicating that they share usage patterns with all groups older than 3;0, while still having a specific characteristic.

The distribution of clusters relative to each other shows the following: the two small clusters 2 and 3 are unique to the two youngest age groups (2;0-3;0), and located at a great distance from other clusters. The three remaining clusters are more adjacent to each other, yet showing distinct locations relative to their usage by specific age groups. Cluster 1 is associated with the three year olds; cluster 4 is associated with the four and five year olds; and cluster 5 is associated with the oldest group of seven year olds.

Discussion

The present paper analyzed the distribution of zero and pronominal subjects in 1st and 2nd person by verb tense, number, person, and utterance role in conversation, in a corpus of peer talk interactions of children aged 2;0-8;0 years. A main, general finding is that zero and pronominal subjects are not evenly distributed relative to the study variables, so that the choice between them cannot be simply regarded as a grammatical alternation. Table 8 shows a summary of these distributions, as follows. 1st person plural subjects in future tense, as well as 1st plural and singular subjects in past tense, are almost always realized as zero; while 1st person singular subjects in future tense occur only as pronouns. In addition, 2nd person plural subjects in both future and past tenses are almost always realized as zero, while 2nd person singular subjects are more evenly distributed. That is, the realization of the subject in Hebrew child language as zero or pronoun is determined by clear linguistic parameters, leaving little room for choice or alternation.

Table 8. Summary of zero and pronominal subject distributions

	1st person		2nd person	
	Singular	Plural	Singular	Plural
Past tense	Zero	Zero	Zero/Pronoun	Zero
Future tense	Pronoun	Zero	Zero/Pronoun	Zero

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

From a developmental perspective, according to the present corpus, most significant differences were found between the youngest group (2;0-2;6) and the rest of the age groups. That is, Hebrew-speaking children in the middle of the third year of life already use zero and pronominal subjects in a systematic way, resembling adult spoken language use (Ariel, 1999). Thus, the youngest children (2;0-2;6) used more zero and fewer pronominal subjects, whereas at 2;6 this distribution was similar to the much older second graders (aged 7;0-8;0); and in the same way, the 2;0-2;6 age group used significantly more plural subjects than the older groups. The person, tense, and utterance type categories portrayed a messier developmental picture, but distribution was shown to be unevenly distributed across the age groups: the 2;0-2;6 and 3;0-4;0 age groups used relatively fewer 2nd person subjects than the other age groups, while the 7;0-8;0 group used more. The 3;0-4;0 age group used fewer past tense verbs and more future tense verbs relative to other age groups. Finally, regarding the distribution of utterance types, we found the following results: (i) the 2;0-2;6 year olds used more Resonating utterances and fewer Response utterances than other groups; (ii) the 2;6-3;0 and 4;0-5;0 age groups used fewer Response utterances than other groups; and (iii) the 7;0-8;0 age group used more Response utterances and fewer Resonance utterances than other groups in the data.

In sum, these findings show that in Hebrew, development starts very early in terms of the parameters affecting the use of zero and pronominal subjects, supporting previous studies on the early use of 1st and 2nd person subjects. Moreover, we show that the acquisition of verb inflectional morphology is crucial for the development of zero and pronominal subjects, and that pragmatic contexts affect subject realization (Armon-Lotem, 2008; Ashkenazi, Gillis, & Ravid, 2020; Ashkenazi, Ravid, & Gillis, 2016; Berman, 1990; Hacohen & Schaeffer, 2007; Schaeffer & Ben Shalom, 2008; Uziel-Karl & Berman, 2000). The distributional results reported in the present paper concern multiple parameters simultaneously, and may form the basis for further hypotheses regarding grammatical development in Hebrew and cross-linguistically.

The present paper took a further step in showing that, in children older than three years, development regarding zero and pronominal subjects in Hebrew concerns changing multifactorial usage patterns. Based on the results of the cluster analysis together with the correspondence analysis, Figure 3 draws the developmental path of Hebrew pronominal subjects.

Figure 3 shows that the earliest detectable usage pattern emerges by 3 years of age, using Initiating utterances, referring to 1st person plural entities (i.e., *we*), in future tense, realized as zero subjects. A similar usage pattern occurred in the 4;0-6;0 groups, but they tended also to refer to 2nd person singular entities (i.e., *you.sg*) in the past. The oldest

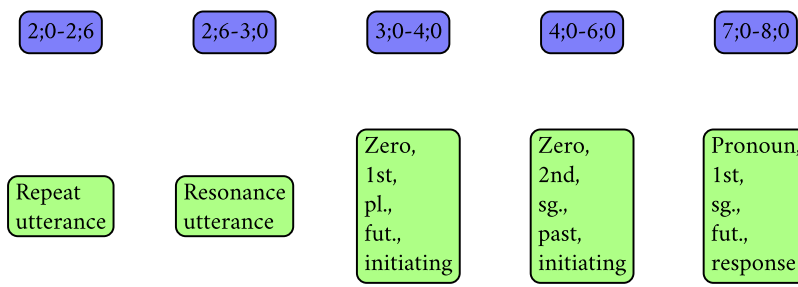

Figure 3. Developmental path of usage patterns of zero and pronominal subjects

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

group is associated with Response utterances, referring to a full 1st person singular pronoun (i.e., *I*) in future tense.

We now present the argumentation for Discourse Profile Constructions of Hebrew zero and pronominal subjects across development in terms of pairing grammatical features of specific usage patterns with their discursive and communicative functions.

From usage patterns to Discourse Profile Constructions: Paring forms with functions

Usage Pattern A (Cluster 1) involves Initiating utterances, referring to 1st person plural entities in future tense, realized as zero subjects, functionally interpreted as a Call for Action. This is a non-narrative mode, mainly used to contemplate and plan while introducing new topics to the conversation in non-subject positions. The 1st person plural subject refers to the peer group as a collective agent, a setting that is best suited for zero reference. The three year olds in our database planned joint actions relatively more than their older and younger peers. Usage pattern A is exemplified in the following:

(18) *hey, na'axil et ha-bubot.*
Hey, we.will.feed ACC the-dolls.
'Hey, we'll feed the dolls.'
(Age group: 3;0-4;0)

(19) *ulay noxal kvar et ha-baflim?*
Maybe we.will.eat now ACC the-waffles?
'Why won't we eat the waffles now?'
(Age group: 4;0-5;0)

Usage Pattern B (Cluster 4) also involved Initiating utterances, this time referring to 2nd person singular entities in the past tense as zero pronouns, mostly used to Comment on the Actions of the Addressee. The use of a singular, 2nd person, zero pronoun subject indicates the accessibility of the referent in the discourse (Ariel, 2001), and the Initiating utterances concern the beginning of a new chunk of conversation. Usage Pattern B mostly occurred in the mid-range age groups (4;0 to 6;0). This usage pattern is exemplified in the following excerpts:

(20) *lo cavata be-kaxol kan!*
NEG you.painted in-blue here!
'You didn't paint blue in here!'
(Age group: 4;0-5;0)

(21) *amit, halo, shaxaxta lehorid la-parpar.*
Amit, hello, you.forgot to.take-off to-the.butterfly.
'Amit, hello, you forgot to take off (the water tattoo) to the butterfly.'
(Age group: 7;0-8;0)

Usage Pattern C (Cluster 5) involved Response utterances referring to 1st person singular pronominal subjects with a future tense verb, used for Planning One's Own Future Actions as a response to some contextual stimulus. Unlike Patterns A and B above, Pattern C concerns responding to utterances as part of the cooperative give and take of

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

conversation, or cooperation in current situations, rather than initiating a discourse chunk. This can be seen in the following example, where speaker B responds to speaker A:

(22) A: *aba amar lexa lalexet lishon.*
Daddy told to.you to.go to.sleep.
'Daddy told you to go to sleep.'
B: *az ani yikax et ha-sapa.*
so I will.take ACC the-couch.
'So I will take the couch.'
(Age group: 7;0-8;0)

This is a more mature usage pattern, that is anchored in previous context, and provides contextual material for conversational development by discussing future actions. Note that 1st person singular in future tense is always realized in our data as a full pronominal subject *I*, raising a question as it is a highly accessible referent – the speaker. However, this usage pattern involves a unique context that calls for a pronoun rather than a zero realization: planning one's own action with regards to the action of others as a point of reference. This context highlights the role of the speaker as opposed to the addressee or relative to some other third-party entity. This is demonstrated in the exchange in (22) as part of a pretend play. The first speaker starts a 'playing chunk', referring to an out-of-context entity as a grammatical subject (*aba* "dad"), and the second speaker responds by referring to his own actions as a reaction to the circumstances that have been laid out for him to play with.

Importantly, it is not the case that 1st person singular is always manifested by a full pronoun; it is only so in the context of future tense. In past tense, where highlighting the self as an agent of the planned actions is not the communicative goal, we may find zero pronouns, as in the final utterance in (23):

(23) A: *Yuli, tistakli ba-xoveret.*
Yuli, look at-the.booklet.
'Yuli, look at the booklet.'
B: *aval ani koshéret.*
But I tie.
'But I'm tying.'
A: *tiksheri lax kol ha-yom!*
You.will.tie to.you all the-day!
'You'll tie for yourself the whole day long!'
B: *kimat siyámti.*
Almost I.done
'I'm almost done.'
(Age group: 5;0-6;0)

Speaker A asks speaker B to look at the booklet, and speaker B is busy doing something else. When referring to her current actions (in the present tense, which is out of the scope of the current paper), speaker B uses a pronoun. However, when referring to herself in a past tense construal, she is using a zero subject ("almost done"). This difference between future and past tense use of zero vs. pronominal 1st person subjects underscores the discursive-functional aspect of subject realization, which is certainly not a matter of a

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

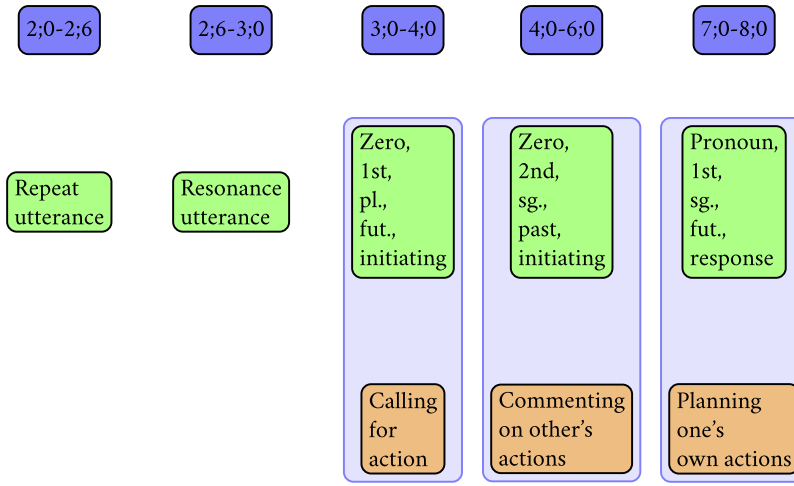

Figure 4. Developmental path of zero and pronominal subjects in Hebrew as Discourse Profile Constructions. Usage patterns of co-occurring formal and conversational features (green boxes) are probabilistically linked to discourse functions (orange boxes), and ordered along a developmental path (purple boxes).

single governing factor. Rather, it is a matter of usage patterns that are anchored in functional contexts, each fitting a different discursive construal and serving a communicative goal. These discourse-based usage patterns are Discourse Profile Constructions (Dattner, 2015, 2019; Dattner et al., 2019). The functions we claim to be paired with the usage patterns as revealed by the cluster analysis are at the discourse level: A – calling for action, B – commenting on the actions of the addressee, and C – planning one's own actions. These are basic discursive scenarios which are linked to multifactorial forms. Illustrating these conclusions, Figure 4 presents an updated version of Figure 3, highlighting the development of Discourse Profile Constructions.

Conclusions

During development, children learn how to use language on various levels such as phonetics, morphology, or syntax. The present paper argues that part and parcel of grammatical development is learning the link between grammar and discourse; that is, learning that particular discursive functions (e.g., the function of "calling for action") are achieved through uttering a particular sequence of words and grammatical categories. Thus, beyond the actual event construal which lies in the lexical semantics of the predicate, or the syntactic-semantic link of argument structure constructions, learning to use language is learning to communicate and to converse.

The current paper shows that while Hebrew zero and pronominal subjects require early acquired knowledge of an inherent part of grammar, children at different ages will not necessarily make use of it in the same way. We also showed that the choice between Hebrew zero and pronominal subjects in 1st and 2nd person can hardly be regarded as an alternation. Rather, each of the possibilities of subject realization is linked to a distinguishable usage pattern. These usage patterns were further shown to correspond with particular discursive and communicative functions, constituting Discourse Profile

https://doi.org/10.1017/S030500092200068X Published online by Cambridge University Press

Constructions with the functions of (A) calling for action, (B) commenting on the action of the other interlocutor, and (C) planning one's own actions. That is, each of these functions is probabilistically served by a particular usage pattern with respect to the person and number of the subject referent, verb tense, the conversational role of the utterance, and the subject realization. We showed that the usage conditions of zero and pronominal subjects change with age, relative to the development of peer conversation. Like previous studies, we have found little development regarding subject grammatical acquisition beyond the age of three years. Our novel addition is in showing that the youngest children in our data (aged 2;0-3;0) have fewer opportunities to use full pronouns compared to their older peers; and that from three years of age the development of pronominal subjects can be best accounted by differences in usage patterns and their associated functions, defined as Discourse Profile Constructions. We conclude that on top of words, and beyond argument structure constructions at the syntactic-semantic level, speakers experience and learn language through basic discursive scenarios. This is of particular importance when working within a usage-based model of language, which assumes that language knowledge stems from language use, and usage naturally changes with age.

References

Alonso-Ovalle, L., **Fernández-Solera, S.**, **Frazier, L.**, & **Clifton, C.** (2002). Null vs. overt pronouns and the topic-focus articulation in Spanish. *Italian Journal of Linguistics*, **14**, 151–170.

Ariel, M. (1999). The development of person agreement markers: From pronouns to higher accessibility markers. In M. Barlow & S. Kemmer (Eds.), *Usage based models of language* (p. 197–260). CSLI.

Ariel, M. (2001). Accessibility theory: An overview. In T. Sanders, J. Schliperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 29–87). Amsterdam & Philadelphia: John Benjamins.

Armon-Lotem, S. (2008). Subject use and the acquisition of verbal agreement in Hebrew. In N. Gagarina & I. Gülzow (Eds.), *The acquisition of verbs and their grammar: The effect of particular languages* (pp. 45–67). Dordrecht: Springer.

Arnon, I. (2010). Rethinking child difficulty: The effect of NP type on children's processing of relative clauses in Hebrew. *Journal of Child Language*, **37**(1), 27–57.

Ashkenazi, O., **Gillis, S.**, & **Ravid, D.** (2020). Input-output relations in the early acquisition of Hebrew verbs. *Journal of Child Language*, **47**, 509–532.

Ashkenazi, O., **Ravid, D.**, & **Gillis, S.** (2016). Breaking into the Hebrew verb system: A learning problem. *First Language*, **36**(5), 505–524.

Avram, L. (2018). Some notes on the study of early subjects in child Romanian. *Bucharest Working Papers in Linguistics*, **21**(1), 5–16.

Barðdal, J. (2004). The semantics of the impersonal construction in icelandic, german and faroese: Beyond thematic roles. In W. Abraham (Ed.), *Focus on germanic typology* (pp. 105–137). Berlin: Akademie Verlag.

Berman, R. A. (1980). The case of an (S)VO language: Subjectless constructions in modern Hebrew. *Language*, **56**(4), 759–776.

Berman, R. A. (1990). On Acquiring an (S) VO Language: Subjectless Sentences in Children's Hebrew. *Linguistics*, **28**(6), 1135–1166.

Berman, R. A. (2011). Revisiting impersonal constructions in Modern Hebrew: Discourse-based perspectives. In A. Malchukov & A. Siewierska (Eds.), *Studies in Language Companion Series* (Vol. **124**, pp. 323–356). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/slcs.124.12ber

Biberauer, T., **Holmberg, A.**, **Roberts, I.**, & **Sheehan, M.** (Eds.). (2009). *Parametric variation: Null subjects in minimalist theory*. Cambridge, UK; New York: Cambridge University Press.

Blevins, J. P. (2013). The information-theoretic turn. *Psihologija*, **46**(4), 355–375.

Borer, H. (1989). Anaphoric AGR. In O. Jaeggli & K. J. Safir (Eds.), *The Null Subject Parameter* (pp. 69–109). Dordrecht: Kluwer.

Borer, H., & **Wexler, K.** (1987). The maturation of syntax. In T. Roeper & E. Williams (Eds.), *Parameter setting* (pp. 123–172). Dordrecht: Springer.

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, **82**(4), 529–551.

Charnavel, I. (2019). Presupposition failure and intended pronominal reference: Person is not so different from gender after all. *Linguistic Inquiry*, **50**(2), 372–387.

Conwell, E. (2020). Argument types affect children's acquisition of a novel argument structure. *PsyArXiv preprint. April*, **13**. doi: 10.31234/osf.io/jy5ba

Dattner, E. (2015). *Mapping Hebrew dative constructions* (Unpublished doctoral dissertation). Tel Aviv University.

Dattner, E. (2019). The Hebrew dative: Usage patterns as discourse profile constructions. *Linguistics*, **57**(5), 1073–1110. doi: 10.1515/ling-2019-0022

Dattner, E., **Kertes, L.**, **Zwilling, R.**, & **Ravid, D.** (2019). Usage patterns in the development of Hebrew grammatical subjects. *Glossa: A journal of general linguistics*, **4**(1), 129. doi: 10.5334/gjgl.928

del Prado Martín, F. M., & **Du Bois, J. W.** (2015). Syntactic alignment is an index of affective alignment: An information-theoretical study of natural dialogue. In *Proceedings of the 37th annual conference of the cognitive science society, San Jose, California.*

Delucchi, K. L. (1993). On the use and misuse of chi-square. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Statistical issues* (pp. 295–320). Lawrence Erlbaum Associates, Inc.

Divjak, D., & **Gries, S. T.** (2006). Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, **2**(1), 23–60.

Dixon, R. M. W. (1994). *Ergativity*. Cambridge: Cambridge University Press.

Dixon, R. M. W. (2010). *Basic linguistic theory*. Oxford; New York: Oxford University Press.

Dowty, D. (1991). *Thematic proto-roles and argument selection*. Language, 547–619.

Du Bois, J. W. (1987). The discourse basis of ergativity. *Language*, **63**, 805–855.

Du Bois, J. W. (2003). Argument structure: Grammar in use. In J. W. Du Bois, L. E. Kump, & W. J. Ashby (Eds.), *Preferred argument structure* (pp. 11–60). Amsterdam & Philadelphia: John Benjamins.

Du Bois, J. W. (2014). Towards a dialogic syntax. *Cognitive Linguistics*, **25**(3), 359–410.

Forrester, M. A., & **Cherrington, S.** (2009). The development of other-related conversational skills: A case study of conversational repair during the early years. *First Language*, **29**(2), 166–191.

Forsythe, H., **Greeson, D.**, & **Schmitt, C.** (2019). Learnability in romance : How indirect input helps children acquire the contrast between null and overt subjects. *Proceedings of the 43rd Boston University Conference on Language Development. Somerville, MA: Cascadilla Press*.

Forsythe, H., **Greeson, D.**, & **Schmitt, C.** (2022). After the null subject parameter: Acquisition of the null-overt contrast in Spanish. *Language Learning and Development*, **18**(2), 171–200.

Friedmann, N., & **Novogrodsky, R.** (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language*, **31**(3), 661–681.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University Of Chicago Press.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, **7**(5), 219–224.

Golinkoff, R. M., & **Hirsh-Pasek, K.** (2008). How toddlers begin to learn verbs. *Trends in Cognitive Sciences*, **12**(10), 397–403.

Gries, S. T. (2010). Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon*, **5**(3), 323–346. doi: 10.1075/ml.5.3.04gri

Grinstead, J. (2016, November). Root Infinitives in Child Language and the Structure of the Clause. In J. L. Lidz, W. Snyder, & J. Pater (Eds.), *The Oxford Handbook of Developmental Linguistics* (Vol. **1**). Oxford University Press. doi: 10.1093/oxfordhb/9780199601264.013.15

Grinstead, J. (2000). Constraints on the computational component vs. grammar in the lexicon: A discussion of Bates & Goodman. *Journal of Child Language*, **27**(3), 737–743.

Gutman, E. (2004). Third person null subjects in Hebrew, Finnish and Rumanian: An accessibility-theoretic account. *Journal of Linguistics*, **40**(3), 463–490. doi: 10.1017/S0022226704002890

Hacohen, A., & **Schaeffer, J.** (2007). Subject realization in early Hebrew/English bilingual acquisition: The role of crosslinguistic influence. *Bilingualism: Language and Cognition*, **10**(3), 333–344.

Halevy, R. (2020). Impersonal and pseudo-impersonal constructions. In R. A. Berman (Ed.), *Usage-Based Studies in Modern Hebrew: Background, Morpho-lexicon, and Syntax* (pp. 539–582). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/slcs.210.16hal

Husson, F., **Josse, J.**, & **Pagès, J.** (2010). Principal component methods – hierarchical clustering – partitional clustering: Why would we need to choose for visualizing data? *Applied Mathematics Department* 17 (2010).

Hyams, N. (1986). *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.

Hyams, N. (2011). Missing subjects in early child language. In J. de Villiers & T. Roeper (Eds.), *Handbook of generative approaches to language acquisition* (pp. 13–52). Dordrecht: Springer.

Hyams, N., & **Wexler, K.** (1993). On the grammatical basis of null subjects in child language. *Linguistic Inquiry*, **24**(3), 421–459.

Keenan, E. L., & **Comrie, B.** (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry*, **8**(1), 63–99.

Kibort, A. (2011). The elephant in the room: The impersonal *-ne/-te* construction in Polish. In A. Malchukov & A. Siewierska (Eds.), *Studies in Language Companion Series* (Vol. **124**, pp. 357–394). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/slcs.124.13kib

Lambrecht, K. (2000). When subjects behave like objects: An analysis of the merging of S and O in sentence-focus constructions across languages. *Studies in Language*, **24**(3), 611–682.

Lê, S., **Josse, J.**, & **Husson, F.** (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, **25**(1), 1–18.

Levy, Y., & **Vainikka, A.** (2000). The development of a mixed null subject system: A cross-linguistic perspective with data on the acquisition of Hebrew. *Language Acquisition*, **8**(4), 363–384.

Li, C. N., & **Thompson, S. A.** (1976). Subject and topic: A new typology of language. In C. N. Li (Ed.), *Subject and topic* (pp. 457–489). New York: Academic Press.

Lustigman, L. (2015). Non-finiteness in early Hebrew verbs. In O. Bat-El (Ed.), *The Acquisition of Hebrew Phonology and Morphology* (pp. 211–229). BRILL. doi: 10.1163/9789004280151_012

Melnik, N. (2007). Extending partial pro-drop in Modern Hebrew: A comprehensive analysis. In S. Möller (Ed.), *Proceedings of the 14th international conference on head-driven phrase structure grammar, Stanford department of linguistics and CSLI's LinGO lab* (pp. 173–193). Stanford, CA: CSLI Publications.

Mohanan, K., & **Mohanan, T.** (1990). Dative subjects in Malayalam: Semantic information in syntax. In M. K. Verma & K. Mohanan (Eds.), *Experiencer subjects in South Asian languages* (pp. 43–58). Stanford, Calif.: CSLI Publications.

Morris, W. C., **Cottrell, G. W.**, & **Elman, J.** (2000). A connectionist simulation of the empirical acquisition of grammatical relations. In S. Wermter & R. Sun (Eds.), *Hybrid neural systems* (pp. 175–193). Berlin, Heidelberg: Springer Berlin Heidelberg.

Orozco, R. (2016). Subject pronoun expression in Mexican Spanish: ¿Qué pasa en Xalapa? *Proceedings of the Linguistic Society of America*, **1**, 7. doi: 10.3765/plsa.v1i0.3703

Orvig, A. S., **Marcos, H.**, **Morgenstern, A.**, **Hassan, R.**, **Leber-Marin, J.**, & **Parès, J.** (2010). Dialogical beginnings of anaphora: The use of third person pronouns before the age of 3. *Journal of Pragmatics*, **42**, 1842–1865.

Paradis, J., & **Navarro, S.** (2003). Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *Journal of Child Language*, **30**(2), 371–393.

Park, M.-K. (2001). Subject-less Clefts in Korean: Towards a Deletion Analysis. *Language Research*, **37**(4), 715–739.

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

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schaeffer, J. C., & **Ben Shalom, D.** (2008). On child subjects in a partially *pro* -drop language. In S. Armon-Lotem, G. Danon, & S. D. Rothstein (Eds.), *Linguistik Aktuell/Linguistics Today* (Vol. **134**, pp. 245–266). Amsterdam: John Benjamins Publishing Company. doi: 10.1075/la.134.10onc

Serratrice, L. (2005). The role of discourse pragmatics in the acquisition of subjects in Italian. *Applied Psycholinguistics*, **26**(03), 437–462.

Serratrice, L., & **Allen, S. E.** (2015). *The acquisition of reference*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Serratrice, L., **Sorace, A.**, & **Paoli, S.** (2004). Crosslinguistic influence at the syntax–pragmatics interface: Subjects and objects in English–Italian bilingual and monolingual acquisition. *Bilingualism: Language and cognition*, **7**(3), 183–205.

Shin, N. L. (2016). Children's Spanish subject pronoun expression: A developmental change. In S. Sessarego & F. Tejedo-Herrero (Eds.), *Spanish language and sociolinguistic analysis* (pp. 155–176). Amsterdam and Philadelphia: John Benjamins Publishing Company.

Shin, N. L., & **Cairns, H. S.** (2009). Subject pronouns in child Spanish and continuity of reference. In *Selected proceedings of the 11th Hispanic linguistics symposium* (pp. 155–164).

Shlonsky, U. (2009). Hebrew as a partial null-subject language. *Studia linguistica*, **63**(1), 133–157.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, **74**(3), 209–253.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA, and London: Harvard University Press.

Uziel-Karl, S., & **Berman, R. A.** (2000). Where's ellipsis? Whether and why there are missing arguments in Hebrew child language. *Linguistics*, **38**(3), 457–482.

Vainikka, A., & **Levy, Y.** (1999). Empty Subjects in Finnish and Hebrew. *Natural Language & Linguistic Theory*, **17**(3), 613–671.

Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition*, **40**(1-2), 21–81.

Veneziano, E., & **Clark, E. V.** (2016). Early verb constructions in French: Adjacency on the left edge. *Journal of Child Language*, **43**(6), 1193–1230.

Wittgenstein, L. (1953). *Philosophical investigations*. New York: McMillan.

Zahler, S. (2018). *The Relationship between Working Memory and Sociolinguistic Variation in First and Second Languages: The Case of Spanish Subject Pronouns*. Unpublished Doctoral dissertation, Indiana University.

Zwiling, R. (2009). *Noun plurals in children's peer talk 2-8 years.* Tel Aviv University MA dissertation.

Cite this article: Dattner E., & Ravid D. (2023). The development of Hebrew zero and pronominal subject realization in the context of first and second person. *Journal of Child Language* 1–27, https://doi.org/10.1017/S030500092200068X