Intonation Units and the Structure of Spontaneous Spoken Language: A View from Hebrew

Shlomo Izre'el

izreel@post.tau.ac.il

Tel-Aviv University

Summary: This paper stems from the initial phases of a research project whose primary goal is the search for a meaningful structural unit of analysis for the study of spoken Hebrew. The end product will include a preliminary analysis of the basic structural units of spontaneous Hebrew in segmental and suprasegmental terms. The working hypothesis for this research is that the Intonation Unit (IU) encapsulates the basic structural unit of the spoken language, with IU Complex (or Utterance) as a higher unit in the hierarchy. In this framework, I will strive to validate or refute accepted, traditional concepts used in the study of spontaneous spoken Hebrew. Preliminary observations of spontaneous spoken Hebrew suggest that neither the clause nor predication carry such basic importance as they do in the analysis of the written language. Therefore, I will check the validity of terms and concepts like sentence, clause, subject and predicate, verbal and non-verbal predication for the syntactic and information-structure analysis of spoken Hebrew. In this paper, I present my working hypothesis, some methodological matters, and preliminary observations.

1. "Old habits die hard"

Until recently, spoken varieties of written languages did not usually draw serious scholarly attention, as they were regarded as degraded forms of the written varieties, consisting of faulty configurations unworthy of description. Exceptions were few, of which the use of spoken French data in the *magnum opus* of Damourette & Pichon (1911-1940) is a prominent example. Interest in the spoken medium and systems of language realized in that medium have arisen in the last decades, notably with the growing availability of large computerized corpora. Linguists have begun to realize that spoken varieties have structure of their own and therefore deserve serious, non-trivial academic attention. As against the earlier view of the written language as superior to the spoken one, 20th century linguistics is said to be inclined to hold the spoken language as primary, and writing as "a means of representing speech in another medium" (Lyons 1968, 38). Hockett (1958, 4) writes: "Old habits die hard. Long after one has learned the suitable technical vocabulary for discussing language directly, rather than via writing, one is still apt to slip. It should afford some consolation to know that it took linguistic scholarship a good many hundreds of years to make just the same transition." I am afraid that even the linguistic community has not yet overcome tradition.

Whereas the study of the oral features of language has a long tradition in itself, it has had the tradition of following the obvious: phonetics and phonology. Morphology as reflected in speech has been given some attention only where the oral medium has implications for the production of words. The special lexicon of vernaculars, notably slang, has also attracted scholarly attention, as it had only minimal manifestations in writing. Syntactic structures of spoken languages have hardly been studied in their own right. Even when linguistics started to show an interest in the linguistic system of speech, syntax seemed to be a stepchild. It still does. Comprehensive treatments of complete syntactic systems of spoken linguistic varieties are few and far between. Still, some significant contributions to the understanding of the syntactic structure of spoken language have

* I thank Eran Cohen and Eitan Grossman for commenting on an earlier draft of this paper.

nevertheless been made. Of these, I would mention Blanche-Benveniste *et al*. 1990; Miller & Weinert 1998; Biber *et al*. 1999; Blanche-Benveniste 1997; Halliday 2004.

Biber *et al*. 1999 is the first large-scale, comprehensive grammar of English that tries to encompass recent advances in corpus linguistics and new ways of understanding the differences between written and spoken varieties and integrate them into a single systematic description. As much as one can admire the results, one cannot ignore the heavy load of traditional linguistics on the authors of this *magnum opus*, as noted by Sinclair (2001, 358): "I have mentioned one or two places where the commentary gets into difficulties because the language obstinately refuses to divide itself into the categories prepared in advance for it."

2. Background: segmental and suprasegmental structure

One of the basic differences between the oral and the written outputs of language is prosody. Although the study of prosody (usually referred to — by way of hyponymic metonymy — as intonation) is centuries old, there are serious gaps in our knowledge of this area. In his conclusion to the second edition of his comprehensive book *Intonation*, Cruttenden (1997, 179) sadly states: "[I]f we consider the full study of intonation to involve sound–meaning correspondences, it is apparent that considerable advances have been made at the phonetic end, but very little at the semantic end." As implied from the list of desiderata and recent achievements that precedes this statement, Cruttenden surely refers also to the syntactic and pragmatic ends.

The growing interest in corpus linguistics and the significant developments in speech technology have enhanced endeavors to better understand the interrelationship between segmental and suprasegmental structure, or between prosody and syntax or discourse structure. This is realized in the relative substantial number of papers devoted to these issues in the recent international two Speech Prosody conferences (Bel & Marlien 2002; 2004). This goes hand in hand with a growing interest in the study of spontaneous, non-elicited speech, of which very little work was done before. Recent research has found that spontaneous speech, notably conversational speech, is very different from other spoken varieties, to the extent that it calls for special research

2.1. The search for a unit for analysis of spoken language: suprasegmental

It has long been recognized that spoken language organizes itself in segments of speech that can be accounted for by their suprasegmental structure. The suprasegmental unit according to which segmentation of the spoken language can be made has been conceived to be dependent mainly on tone, or intonation, and has therefore been termed 'tone group', 'intonation group', 'tone unit', 'intonation(al) phrase', 'intonation unit', or the like. (e.g., Beckman & Pierrehumbert 1986; Halliday 1989; Selkirk 1984; Chafe 1994; Cruttenden 1997; Brazil 1997; Hirst & Di Cristo 1998; Halliday 2004), where the identified prosodic stretch may be identical or different in some respects among the various approaches. Different paths have been used to explicate the concept. Whatever approach is taken, it seems that there is a wide consensus that the intonation unit encapsulates a functional, coherent segmental unit, be it syntactic, semantic, informational, or the like (see below).

It seems commonly accepted that an intonation unit is a coherent intonation contour, and some would define the intonation unit in these terms (Chafe 1994; Du Bois *et al*. 1992; 1993; Tao 1996; etc.). "A coherent intonation contour", while rather easily perceivable, is very hard to define in itself by acoustic, formal terms, nor is it easy to define an intonation unit by any other internal criteria (Cruttenden 1997). In practice, segmentation of a discourse flow into intonation units is made by detecting their boundaries, whereas internal criteria are brought into consideration only secondarily (Cruttenden 1997). This practice has been used successfully in transcribing large corpora (Du Bois *et al*. 1992; 1993; Du Bois 2004; Cresti & Moneglia 2004; cf. also Cheng, Greaves & Warren 2005, following the methodology of Brazil 1997). Theory has also inclined

towards the delimitation of the intonation unit — or 'intonational phrase' — by reference to 'boundary tones': "Each *intonational phrase* provides an opportunity for a new choice of tune, and ... some parts of the tune serve to mark the *phrase boundaries*" (Pierrehumbert & Hirschberg 1990, 272); "Rappelons que le rapport de dominance dépend uniquement des tones finals; il est insensible aux éléments intonatifs apparaissant ailleurs dans le groupe" (Blanche-Benveniste *et. al.* 1990, 172). Other theories tend to relate to the hierarchical structure of prosody, describing larger (higher) units as combinations of smaller (lower) ones (e.g., Nespor & Vogel 1986; Ladd 1986). Some of these studies, based on phonological theory, still use syntactic constructs for defining prosodic units, although — as emphasized by Nespor & Vogel — the syntactic and prosodic units do not necessarily overlap. For recent theories of prosody see Ladd 1996; Gussenhoven 2004.

The major perceptual and acoustic cues for boundary recognition are the following: (1) final lengthening; (2) initial rush; (3) pitch reset; (4) pause (Cruttenden 1997; Du Bois *et al*. 1992; Hirst & Di Cristo 1998; for Hebrew see Amir, Silber-Varod & Izre'el 2004; Silber-Varod 2005). The final tone of an IU carries with it functional load in terms of discourse structure and information structure, with implications for syntax. Thus, for American English, three major categories of transitional continuity have been suggested, mainly suggesting interaction between the interlocutors: terminal, continuing and appeal. The first two mark final or non-final contour respectively; the third is, *mutatis mutandis*, equivalent to "yes-no" questions (Du Bois *et al.* 1992; for other systems see, *inter alia*, Hirst & Di Cristo 1998).

Given the problems of defining a coherent intonation contour, and the additional prosodic cues for defining the unit boundaries, I tend to prefer the term 'prosodic unit' over 'intonation unit'.[1] However, for the sake of consistency with other studies discussed below, I will adhere, at this point, to the accepted term 'intonation unit' (henceforth: IU). Either term should be confined to the phonological, suprasegmental features of the stretch, and another term, denoting the structural, segmental stretch confined by prosodic boundaries, should be sought.

2.2. The search for a unit for analysis of spoken language: segmental

For Generative Linguistics, the 'sentence' is a central construct. Therefore, many studies stemming from this school and its offshoots tend to discuss prosodic structure by reference to the sentence (see, e.g., Selkirk 1984; Steedman 2001). For other schools, the sentence is not a valid structural unit for the description and analysis of spoken language, notably its spontaneous, conversational varieties.

Halliday (e.g., 1989; 2004), whose work has had a significant impact on wide linguistic circles, promotes the use of the clause as the basic unit of the spoken language. Whereas Halliday's research was based on English, Miller & Weinert (1998) applied this insight over a body of three languages, including Russian and German, in addition to English. Their work concentrates on spontaneous spoken language. Similarly, Biber *et al.* (1999), in a special chapter devoted to conversational English, look at what they term C-units, i.e., "clausal and non-clausal units". The latter are special cases of discourse, which, as it stands, do not fit into the conventional syntactic point of reference.

The work of GARS (Groupe Aixois de Recherche en Syntaxe) at the Université of Provence, which contributes to the study of spoken French, preceded this endeavor. It further pinpointed the differences between the spoken and the written language. "La syntaxe de la phrase et des propositions, fondée sur les catégories grammaticales et leurs fonctions, ne suffit pas à rendre compte de certaines organisations de la langue parlée" (Blanche-Benveniste 1997, 111). Therefore, GARS distinguishes between *syntaxe* and *macro-syntaxe*; the latter is designed to deal with

[1] I thank Cyril Auran for suggesting the use of the term 'prosodic unit'.

structures that can be either equivalent to or larger than the traditional 'sentence' or 'clause'. The relations between macro-syntactic elements are different from traditional grammatical relations. In many ways, macro-syntax of this French school is treated within the framework of what would be considered pragmatics or discourse analysis in other schools. For others, among which the functional-systemic school holds a prominent position, the closest equivalent to the conceptual framework of macro-syntax would be 'information structure' (e.g., Halliday 2004; also, *inter alia*, Lambrecht 1994). For Halliday, there is also an intermediary analysis of the clause, where the order of components makes the basis of analysis.

Some of the studies mentioned above do not base their analyses on prosodic structure (notably Biber *et al*. 1999; also Miller & Weinert 1998; Lambrecht 1994). For others, prosody is a basic component of the spoken language, to the extent that any analysis of its structural units would stem from prosodic parsing. Thus Brazil (1995), leaning on prosody, takes a different route in analyzing the language of spoken narrative. Brazil rejects the validity of what he calls "sentence grammar", and suggests a unit that he calls "increment", the basic unit of the syntax of targeted, purposeful, linear oral communication.

Chafe's analyses of spoken language (culminating in Chafe 1994) are based on his cognitive theory of information flow. Chafe regards IU as the basic unit of discourse. He describes the relationship between IU and clauses as follows: "Each clause verbalizes the idea of an event or state, and usually each intonation unit verbalizes a different event or state from the preceding" (Chafe 1994, 69). Chafe distinguishes between fragmentary, substantive, and regulatory IUs: Fragmentary IUs are those that have not come to a successful end; substantive IUs are those that convey ideas and events, states or referents; regulatory IUs are those whose function is the regulation of information flow or the interaction between the speakers.

With Chafe's finding, i.e., that clauses and substantive IUs coincide by 60% (Chafe 1994), the question was raised concerning the centrality of the clause in the study of spoken language. Several studies took a similar path of exploration, basing on the premise that the IU is the basic unit of conversational language, checking the constituency of IUs in terms of clauses and NPs and mapping clauses into IUs. These studies covered a range of languages: English; Mandarin Chinese, Japanese, Finish, Wardaman (Australian), Sasak (Western Astronesian) (Iwasaki & Tao 1993; Croft 1995; Tao 1996; Helasvuo 2001; Matsumoto 2003; Croft 2005; Wouk ms). It seems that most authors are happy with the notion of clause as a valid category for the study of spontaneous, conversational language (cf. also Hirst & Di Cristo 1998), either in its most expanded structure or in a reduced form, such as the predicate and its core arguments (Croft 1995; 2005; Helasvuo 2001; cf. Du Bois 1987 for the strategy of preferred argument structure). On the other hand, Tao has found that Mandarin consists of a relatively large number of IUs consisting of NPs (cf. Iwasaki & Tao 1993 for both Mandarin and Japanese; but — with different conclusions for Japanese — Matsumoto 2003). To conform to these findings, Tao suggests the concept of 'speech unit', "the correspondence of grammatical elements and intonation units, which is real for language production and analytically advantageous for the study of language" (Tao 1996, 175). Suggesting a dynamic grammar that operates at the level of spontaneous natural speech, Tao further questions the usefulness of the 'clause' for the analysis of spoken language.

I have mentioned above transcriptional strategies of large corpora. The use of prosodic parsing in the C-ORAL-ROM corpus of Romance languages resulted in significant implications for the understanding of the structure of the languages comprising this corpus. The basic structural unit of spoken language is suggested to be the 'utterance', which is defined operatively and theoretically: "The operative definition of the utterance is such that every expression marked by a prosodic terminal break is an utterance. ... From our theoretical point of view, an utterance corresponds to the accomplishment of a speech act, as defined by Austin (1962)" (Cresti 2004, 210). One must note at this point, that an 'utterance' in the Romance corpus equals a set of any

number (including one) of sequential intonation units as defined by Chafe and his followers, i.e., all IUs between two final boundaries (=terminal breaks). Therefore, any statistics presented for the C-ORAL-ROM is utterance-based rather than based on IU count. This means that there are many more IUs in the C-ORAL-ROM that lack predication than the number of 30% suggests. For the C-ORAL-ROM compilers, "[s]ince the average 30% of utterances are verbless, all the definitions based on clause structure and verbal predication appear to be inadequate for spoken language analysis purposes" (Moneglia 2004).

3. Prosody, syntax and information structure in the study of Spoken Israeli Hebrew

Whereas phonetics and phonology of IH, being a prominent feature of the spoken medium, have been dealt with in a number of significant studies, work on prosody has been extremely scanty, and can be ascribed mainly to a single scholar, viz., Laufer (1987). Laufer's work strives for comprehensiveness, suggesting a set of intonation patterns ('tunes') and includes observations on IU boundary cues. However, his studies are based on elicited speech, and his examples are far from reflecting spontaneous, colloquial speech.

Looking at the extant research of Israeli Hebrew, there is an embarrassingly small number of comprehensive treatments of its syntax, and there are even fewer treatments of the syntax of spoken Hebrew (Waldman 1989; Khan 2003; to the studies listed there, one should add especially Schwarzwald 2001). Most studies of spoken Hebrew syntax are limited in scope, taking for their subject of observation specific structures or, more commonly, individual features; the same applies to studies in pragmatics or discourse analysis (Waldman 1989; Khan 2003). A notable exception is Shatil 1994 on syntactic patterns, which follows the methodology of Rosén (1966) and describes patterns that were not listed by Rosén. Shatil's study is worthy of mentioning also because he based his study not on sporadic elicitation of textual material, but on a coherent corpus. Another study worthy of mentioning is Yatziv-Malibert 2002, which, although limited in scope to some syntactic patterns of the spoken language, has paved a methodological path for syntactic observations without precedent in the study of spoken Hebrew. Yatziv-Malibert also used a coherent spoken corpus for her study. The only comprehensive study of Hebrew dealing with Information structure *per se* I know of is Kuzar's dissertation (1989) on information structure in Israeli Hebrew. Kuzar studied (non-spontaneous) spoken and written texts.

Indeed, the study of spoken Hebrew is still in swaddling clothes. Without a corpus of data, there cannot be any research. Random data retrieval enables observation and insights. Authentic research that stems from research hypotheses and includes an examination of the data, followed either by corroboration or by refutation of the hypotheses (not only theory for its own sake), is possible only in a limited way when the research has no proper database. The study of IUs reported here stems from the preparatory stage taken towards the compilation of the Corpus of Spoken Israeli Hebrew (*CoSIH*) (see below, §5.3.1).

Since the spoken medium is acoustic, linear and temporally extended, visual transmission is necessary in order to enable any research, except, perhaps, for such as focused on individual, small units. Even in this latter case, one needs to transform sound into the visual medium in order to publish the results. The linguist must therefore use a transcript of the spoken text. Any type of transcription, including the narrowest one, is based on theory, since there is no way of transforming the infinite range of acoustic features into phonetic symbols. Therefore, any type of transcription must be anchored in theory. The theoretical ground depends on research goals (Ochs 1979; Du Bois 1991; Crowdy 1994; Kennedy 1998; Blanche-Benveniste 2000).

Among the first steps taken in the preparatory stages of the planning of *CoSIH* was the decision regarding the type of transcription to be used. Apart from samples in narrow phonetic

transcription, the bulk of *CoSIH* will be transcribed in the Hebrew standard orthography, whereas the transcription is meant to be aligned with the sound when *CoSIH* is disseminated (Izre'el 2005). According to the theoretical stand of the *CoSIH* team, the texts in their visual form must reflect the nature of the spoken language as consisting of IUs. This decision follows the Chafe–Du Bois methodology of transcription (Du Bois *et al*. 1992; 1993). This method had been used before for transcribing spoken Israeli Hebrew (e.g., Cahana-Amitay & Katsenberger 1999; Maschler 2001), and has already been tried on spontaneous texts elicited for the pilot study of *CoSIH* (Izre'el & Rahav 2004) and already during *CoSIH*'s preparatory stage (Izre'el 2002).

These recorded data have already been used for two major studies on spontaneous spoken Hebrew, based on the IU theorem: Smadar Cohen's pioneering research on the prosodic handling of questions (Cohen 2004) and Vered Silber-Varod's study of IU boundaries (Silber-Varod 2005; see already Amir, Silber-Varod & Izre'el 2004). Still, the syntactic aspect of what constitutes an IU in Hebrew has never come up, to the best of my knowledge.

4. Spoken Israeli Hebrew: some pertinent basic notions

Hebrew belongs to the Semitic family, which in itself is a branch of the larger Hamito-Semitic (or Afro-Asiatic) phylum. Like other languages in this phylum, Hebrew presents two major features that are pertinent to our discussion here: Firstly, clauses can be either verbal or non-verbal. A non-verbal predicate can consist, in principle, of any part of speech; e.g., a noun (ex. 1), an adjective (ex. 2), a prepositional phrase (ex. 3), a clause (ex. 4), etc.[2]

(1) *hakol shvilim* |
the-all paths
'All (roads) are unpaved roads.' (Conversation 2, 251)[3]

(2) *ha-kvish ha-ze patuax* ||
the-road the-this open
'This road is clear.' (Conversation 1, 344)

(3) *ani be-kurs* ||
1SG in-course
'I am attending a course.' (Conversation 1, 714)

(4) *ha-she'ela ma ze naxon* ||
the-question what this right
'The question is what is right.' (Conversation 2, 125)

One should note at this juncture, that word order is not strict, and varies according to syntactic and pragmatic constraints. Needless to say, prosody also plays a role in the structure of an utterance.

Secondly, a verb always constitutes a clause in its own right. In other words, a verb is a morphological manifestation of both a subject and its predicate (ex. 5). It can further include a pronominal complement (ex. 6).

[2] Transcription notations: each IU is transcribed in a separate line; speaker marking (a:); final tones: terminal (||); continuing (|), appeal (/); pauses: short (..), medium (...), long (.....), extra-long (..5..), noting length in seconds; truncated IU (—); truncated word (_); non-verbal sounds (<creak>) (cf. Izre'el 2002, following in essence Du Bois *et al*. 1992; 1993). Other notations: morphemic boundary (-); clitic boundary (=).

[3] For the textual references see below, §6.1.

(5) *balaa-ti*
swallowed-1SG
'I swallowed-up' (Conversation 1, 150)

(6) *azav-ti-ha*
left-1SG-her
'I left her'

Whereas the latter construction is quite rare in spoken Israeli Hebrew (and therefore not attested in any of the three conversations that form the corpus for this paper), there is a tendency in spoken Israeli Hebrew for the cliticization of prepositional phrases that include a complement pronoun:

(7) *lakax-nu=ot-a*
took-1PL=ACC-her
'we took it' (Conversation 2, 69)

(8) *her'e-ti=(l)xa*
show-1SG-to-you(SGM)
'I-showed you' (Conversation 2, 615)

One should note, however, that in spoken Israeli Hebrew, when not directly following the verb, these complements appear as independent forms; this phenomenon is general in written Hebrew, where no such cliticization is observed.

Lastly, Hebrew has across-the-board agreement: person, gender and number for personal pronouns and finite verbs, gender and number for the nominal domain (substantives and adjectives).

5. The present study: goal, working hypothesis, methods

5.1. Goal

This paper stems from the initial phases of research whose main goal is to look for a meaningful structural unit of analysis for the study of spoken Hebrew. The end product will include a preliminary analysis of the basic structural units of spontaneous Hebrew in segmental and suprasegmental terms.

5.2. Working hypothesis and conceptual framework

The working hypothesis for this research is that the Intonation Unit (IU) encapsulates the basic structural unit of spontaneous spoken Israeli Hebrew. I will further look for the next unit in the hierarchy, which I hypothesize to be IU complex or Utterance (see Cresti & Moneglia 2004 above).

In this framework, I will strive to validate or refute accepted, traditional concepts of the study of spontaneous spoken Hebrew. I will check the validity of terms and concepts like sentence, clause, subject and predicate, and verbal and non-verbal predication for the syntactic and informational analysis of spoken Hebrew. As will be shown below, it seems that neither the clause nor predication carry such basic importance as they do in the written language. Either an IU or an IU complex may consist of a traditional clause, but it need not be so, so that the clause may be unwarranted as the basic unit of analysis.

By choosing this methodology, I choose the side of the hearer as my point of departure. Since the parsing of speech flow is made perceptually by prosodic boundaries, the starting point for my structural analysis is prosody. Another possible way is to start from the side of the speaker, and decide upon the basic structural unit of the spoken language and see how it is organized in prosodic

units. This will be done by mapping clauses (for my working definition of a 'clause' see below, §5.3.2) onto IUs, as has been done in previous research (see above, §2.2). This point of departure will be used to counterbalance results obtained by the first method.

A note on terminology and conceptual frames: The concepts of 'subject' and 'predicate' are deeply rooted in Western grammatical tradition, and one can hardly find opponents to their suitability for analyzing languages (cf. Jespersen 1924; Lyons 1968; Matthews 1981; Crystal 1997; Biber *et al*. 1999). When one comes from the study of Semitic languages or their like, the very existence of sentences or clauses with no verb may raise doubt about ways to handle the relationship between the components of such clauses. Indeed, voices have been raised against the wisdom in adhering to these terms in their accepted conceptual framework, not only for Semitic and related languages, but also for "western" ones (e.g., Polotsky 1944; 1962; Shisha-Halevy 1986; Kibrik 2001; Izre'el & Cohen 2004).

5.3. Methods

5.3.1. The corpus for this study

As mentioned above, the study of IUs reported here stems from the preparatory stage taken towards the compilation of the Corpus of Spoken Israeli Hebrew (*CoSIH*). *CoSIH* is designed to include a representation of most varieties of spoken Hebrew as it is used in Israel today. For analytical purposes, it will use a conceptual tool in the form of a multidimensional matrix combining demographic and contextual tiers. Daylong, non-stop recordings of randomly selected informants will be collected with respective sociolinguistic data. These recordings will be evaluated, and a sample from each will be transcribed, to set up a five-million-word corpus. *CoSIH* is, to the best of my knowledge, the first corpus designed to integrate both demographic and contextual criteria in its compilation of texts. For a detailed description of the *CoSIH* project see Izre'el, Hary and Rahav 2001.

CoSIH's pilot consists of recordings made by about four dozen recorded datasets, each consisting of 8 to 16 hours of longitudinal recordings. The recording-recorded informants include people from a variety of backgrounds: ethno-cultural, age, sex and education (Izre'el & Rahav 2004). I have adopted this schema for the compilation of the corpus for the proposed research, so as to include a variety of texts by diverse people and from a variety of contextual situations, in order to obtain a fair representation of text types. As against the conceptual framework of *CoSIH*, whereby native and non-native speakers are treated alike for the compilation of the corpus, I believe that for conducting the research advanced here, a more coherent group of speakers is preferable. Therefore, the corpus for this study includes only native speakers of Hebrew, altogether 20 informants. Of course, each of these datasets includes speech not only of the recording informants but also of a variety of interlocutors. I estimate the total number of speakers for this research to be more than 50. Given the results of previous research on this topic, 500 IUs of each dataset will form a good sample for analyses, both quantitative and qualitative. Therefore, the total number of IUs that will form the corpus of this study is expected to be about 10,000.

5.3.2. Research schema

Along the path paved by previous studies, I use both qualitative and quantitative methods. A detailed database is used to mark the following features of each IU: (1) The function of its final tone: terminal (||), continuing (|) or appeal (/); (2) Whether the IU is fragmentary, substantive or regulatory;[4] (3) Whether it includes a clause; (4) Whether an IU includes more than a single clause;

[4] It may well occur that an IU consist of both substantive and regulatory elements (cf. Matsumoto 2003). In such cases, the IU will be classified as substantive.

(5) If an IU consists or includes a clause (or more than one), whether the clause (or one of them) can be regarded formally as an embedded (or subordinate) clause; (6) In clausal IUs, what elements of the clause it includes beside the predicate; (7) For a non-clausal IU, whether it can be regarded as consisting of a part of a clause (one of its core arguments or any other functional element, with a specification of this element and its position as related to the predicate); (8) In clausal IUs, what type of predicate it includes (verbal, nominal, etc.); (9) For verbal clauses, what type of verb they consist of from both the syntactic and semantic aspects (e.g., its valence, verb introducing direct speech, verb of perception). Wherever applicable, for each IU, the accented (prominent) syllables are marked in the transcription, to enable an alternative analysis of the data at a later stage. In addition, each dataset is supplied with annotation regarding the speakers and the type of speech recorded, for further exploitation of the data regarding demographic, and especially, contextual variation (cf. Cresti & Moneglia 2004).

The notion of 'clause' can be defined in various ways, notably with regard to its constituents. A common approach would define 'clause' as a construct including a (verbal) predicate together with its arguments and peripheral elements (see, e.g., Van Valin & LaPolla 1997; Givón 2001). The study of Hebrew takes a more traditional approach, and defines a 'clause' simply as consisting of a N(oun)P(hrase) and a P(redicate)P(hrase) (Rubinstein 1968; cf. Rosén 1977; Coffin & Bolozky 2004). The minimal requirement is, therefore, that a clause will consist of NP as subject and of a predicate, which does not need to be a verb (Rodrigue-Schwarzwald & Sokoloff 1992). It is with this minimal requirement that I have decided to tag clauses as components of IUs in my data. In this framework, no arguments are required to be present in the same IU as the predicate. Whereas a verb thus consists of a clause by its morphological constituents (see §4 above and further below), a non-verbal component is regarded as a predicate only if there is an overt subject related to it, either within the same IU (ex. 9) or adjacent to it (ex. 10):

(9) *ani be-kurs* ||
1SG in-course
'I am attending a course.' (Conversation 1, 714)

(10) *at* |
2SGF
.. *memal'-a bevakasha* |
fill- SGF please
'You should fill it, please.' (Conversation 3, 261-2)

In contrast, there are many instances where a predicative tagging was not possible, as in the following example:

(11) *pesel exad anak* |
statue one huge
'one huge statue,' (Conversation 2, 721; for the context see below, §6.2)

We have seen above that a Hebrew verb is a synthetic form, consisting of both a subject and a predicate. As such, it can be regarded as consisting of a clause in itself (Goldenberg 1998), to which a complemental element can also be annexed, as exemplified above. Of course, an overt nominal referent ("subject") can form part of the clause. Furthermore, a pronominal referent ("subject") can sometimes precede the verb as well. The reasons for this type of construction are manifold, and can be explained, *inter alia*, by morphophonological and by pragmatic constraints (cf., e.g., Ariel 2000). On one hand, it can be regarded as the "real" subject, with the verb presenting an agreement element. On the other hand, the internal marker can be regarded as the subject, whereas the external form seen as extrapositive. For Taqbaylit Berber, which presents constructions similar to the Hebrew ones, Mettouchi (forthcoming a, forthcoming b) suggests that only the morphological

marker within the verbal form can be regarded as subject, while the independent pronoun (or equivalent noun) must be regarded as carrying pragmatic functions.[5] The following pair of examples suggests a similar tendency:

(12) *axarei ze* |
after this
xazarti le-Beijing |
returned-1SG to-Beijing
'Then I returned to Beijing,' (Conversation 2, 941)

(13) *be-yangshu* |
in-Yangshuo
eldad lakax-ø rakev_ otobus le-po |
Eldad took-3SGM trai(n) bus to-here
le-Hong Kong |
to-Hong Kong
ve-ani xazarti le-Beijing ||
and-1SG returned-1SG to-Beijing
'In Yangshuo, Eldad took a trai(n) a bus to here, to Hong Kong, and I returned to Beijing.'
(Conversation 2, 902-4)

Whereas in ex. 12 only the verb carries marking of the 1SG, there is an independent 1SG pronoun external to the verb complex in ex. 13, which can be explained by the change of topic.

This approach has some interesting similarities to the '*approche pronominale*' taken by Blanche-Benveniste and GARS for the study of French (Blanche-Benveniste 1975; Blanche-Benveniste *et al*. 1984; 1990). Indeed, the 'pronominal approach' has been adopted by Yatziv-Malibert's for her analysis of spoken Hebrew: "D'après cette approche, la construction verbale ou l'unité prédicative et non pas la phrase est retenue, comme l'unité nasique de description de l'énoncé » (Yatziv-Malibert 2002, 53).

The pronominal approach has been taken to be the basic reference for the analysis of the verbal utterance in the C-ORAL-ROM project (Cresti in Cresti & Moneglia 2004, §6.4).

A thorough research into the distribution and functions of subjects and their syntactic and pragmatic equivalents must therefore be made in order to substantiate the mapping of clauses.

6. Preliminary observations

Since the research is only in its very initial phases, in the following I will present a few observations from the preliminary analyses made so far: first quantitative, then qualitative. These observations will serve only to illustrate the type of questions to be addressed as well as some preliminary impressions from the data at hand.

6.1. Quantitative observations

I will present quantitative accounts of three samples from the data, all recorded by a single informant, a 26 years old male with high-school education, who works in a high-tech company, and

[5] On other languages with synthetic verbs in the context discussed here, yet with different interpretation, see Chafe 1994, ch. 12 on Seneca (Iroquoian) and Croft 2005 on Wardaman (Australian).

he alone was aware of the recording (Izre'el & Rahav 2004). The three datasets are samples from recordings in three different situations:

(1) A conversation in a car during a drive back home from a wedding. The interlocutors of our recording-recorded informant in this case are two colleagues from work, both women about the same age as our informant. The topics of discussion are personal, issues related to personal relations at work, and driving matters. (Referred to as 'Coversation 1'.)

(2) A conversation between the informant and his 51-year-old father, who has an academic education. The topic of discussion is our informant's visit to China and Mongolia, and the conversation consists mainly of descriptions of this trip. (Referred to as 'Coversation 2'.)

(3) A meeting regarding human resources at the high-tech company. The interlocutors are four colleagues, two males and two females, with high-school or academic education, all four in their late twenties or thirties. (Referred to as 'Coversation 3'.)

All recorded people are native speakers of Hebrew. The recordings took place between September and December 2000.[6]

The counts have been designed to suggest the validity of the IU as a basic structural unit of spontaneous spoken Hebrew. In this case, there was no attempt to suggest the mapping of clauses onto IUs, but rather, the goal was to study the nature of occurring IUs in terms of clausal components. Therefore, an IU consisting of an embedded clause is counted as clausal, although for the mapping task such clauses would be regarded as NPs (Tao 1996; Wouk ms). Note that this latter procedure suggests a higher number of clauses than the alternative one. Furthermore, the counts take a simplified approach in comparing the data, e.g., they do not distinguish (the few cases of) IUs consisting of two clauses from those consisting of only one.

	Total	Fragmentary	Substantive	Regulatory	**Total non-fragmentary**
Total	**780 (100%)**	**32 (4.1%)**	**688 (88.2%)**	**60 (7.7%)**	**748 (95.9%)**
Non-clausal		100%	**381 (55.4%)**	**47 (78.3%)**	**428 (57.2%)**
Total clausal			**307 (44.6%)**	**13 (21.7%)**	**320 (42.8%)**
Clausal verbal		100%	169 (55.0%)	6 (46.2%)	175 (54.7%)
Clausal non-Verbal			138 (45.0%)	7 (53.8%)	145 (45.3%)

Table 1: Distribution of IU types: Conversation 1

[6] The *CoSIH* preparatory stage and the pilot study were financed mainly by grants from Tel-Aviv University, with some support from Emory University.

	Total	Fragmentary	Substantive	Regulatory	**Total non-fragmentary**
Total	**1098 (100%)**	**128 (11.7%)**	**870 (79.2%)**	**100 (9.1%)**	**970 (88.3%)**
Non-clausal		100%	**475 (54.6%)**	**90 (90%)**	**565 (58.2%)**
Total clausal			**395 (45.4%)**	**10 (10%)**	**405 (41.8%)**
Clausal non-Verbal		100%	272 (68.9%)	9 (90%)	281 (69.4%)
Clausal verbal			123 (31.1%)	1 (10%)	124 (30.6%)

Table 2: Distribution of IU types: Conversation 2

	Total	Fragmentary	Substantive	Regulatory	**Total non-fragmentary**
Total	**384 (100%)**	**15 (3.9%)**	**309 (80.5%)**	**60 (15.6%)**	**369 (96.1%)**
Non-clausal		100%	**146 (47.2%)**	**51 (85.0%)**	**197 (53.4%)**
Total clausal			**163 (52.8%)**	**9 (15.0%)**	**172 (46.6%)**
Clausal verbal		100%	37 (22.7%)	5 (55.6%)	42 (24.4%)
Clausal non-Verbal			126 (77.3%)	4 (44.4%)	130 (75.6%)

Table 3: Distribution of IU types: Conversation 3

The numbers speak for themselves, and show small percentages of clausal IUs, especially when compared to figures for other languages, notably English. A comparison between the three tables suggests differences in scope of clausal use between the three types of speech represented by the respective three datasets, notably Conversation 3 (meeting). An even more interesting observation is the significantly higher number of verbs attested in Conversation 1 (car drive). Both observations suggest interesting different linguistic behavior in different contextual settings, which certainly deserves a thorough investigation. One may also note the higher number of regulatory IUs in Conversation 3 (meeting) and the high number of fragmentary IUs in Conversation 2 (conversation between a son and his father, which is narrative-descriptional in nature).

6.2. Qualitative observations

In addition to the examples adduced above with qualitative observations, I would like to give here a few examples to share with my audience my impressions regarding the nature of the data at hand, which exemplify substantive non-clausal IUs in the three conversation. The IUs in question, marked in boldface characters, are better described in terms of information structure rather than in terms of predication:[7]

[7] For each IU, its basic type is marked: substantive (S); regulatory (R); fragmentary (F).

(14) [1s] *etmol asi-ti tipul l-a-oto* |
yesterday did-1SG treatment to-the-car
'Yesterday I serviced the car;'

[2s] ***elef shlosh me-ot shekel*** ||
thousand three hunderd-PL Shekel
'(it cost) 1300 shekels.' (Conversation 1, 34-35)

(15) [1s] *yesh sham park* |
EXT there park
'There is a park over there,'

[2r] ..(2.2).. <creak> *lo yodea ma* ||
NEG know what
'I'm not sure.'

[3s] ***kama dunam-im tov-im*** |
some acres-PL good-PL
'(The size of it is) a good number of acres,'

[4s] ***male male gumx-ot*** |
full full alcove-PL
'(it has) many many alcoves,'

[5s] ***im male .. psal-im ktan-im be-alaf-im*** |
with full statue-PL small-PL in-thousand-PL
'with many .. small statues by the thousand,'

[6s] ***pesel exad anak*** |
statue one huge
'(there was) one huge statue,'

[7f] .. ***civ_*** —
.. col(orful)
'(there was another one) col-'

[8s] ***cavua*** |
'colored,'

[9s] ***lo cavua*** |
NEG colored
'(still another) not colored,'

[10s] *hakol bud-ot* |
the-all Budha-PL
'all (these are) Budha(-statue)s,' (Conversation 2, 716-725)

(16) [1s] *atem hayom roc-im la-daat briut irgunit* ||
2PL today want-PL to-know health organizational
'Today you want to become informed about organizational health.'

[2r] *naxon* /
right
'Right?'

[3s] *ze ma she-atem mexaps-im* ||
this what NOMINALIZER-2PL search-PL
'This is what you are looking for.'

[4r] *naxon* /
right
'Right?'

[5s] ***briut shel ha*** |
health of the
'(You are after the) health of the ...'

[6s] ***shel ha-sviut=racon shel-i*** ||
of the-satisfation of-me
'of my satisfaction.'

[7s] ***betor oved*** ||
as worker
'As a worker.'

[8r] *naxon* /
right
'Right?' (Conversation 3, 130-138)

All marked IUs are rhemes (or, in a different terminology, 'comment', 'new'), presenting new information, where no thematic elements are explicitly referred to. In ex. 14, the price of the service is presented without any formal reference to the mentioning of the service in the preceding IU. In ex. 15, the marked IUs also present new information, one at a time. Thematic anchors exist only indirectly in units 3 (the park), 5 (alcoves), 7-9 (various statues), whereas the units 4 and 6 only new information exists with no thematic ('given') anchors present (except for the very idea of the location, viz., the park). In ex. 16, units 5 and 6 together form a rheme, the new information. Its thematic anchor can be found in unit 3, and can even be syntactically related to the cataphoric pronoun *ze* "this", which in itself is an anaphora to 'organizational health' in unit 1. However, the syntactic gap between unit 3 and units 5-6, the resumptive function of the phrasal complex in these two units (5-6), and as well as the prosodic structure of the entire passage, prefers a non-syntactic analysis. As for the rhematic unit 7, it can be regarded an afterthought, as suggested by the terminal tone of the preceding IU.

7. Conclusion

Looking for a meaningful structural unit of analysis for the study of spoken Hebrew, a working hypothesis suggests that it be encapsulated by the Intonation Unit (IU), with IU Complex (or Utterance) as a higher unit in the hierarchy. In this framework, accepted, traditional concepts for the study of spontaneous spoken Hebrew will be challenged. Preliminary observations of spontaneous spoken Hebrew suggest that neither the clause nor predication carry such basic importance as they do in the analysis of the written language. The following quotes may suggest the frame of reference for this study:

> ... il ne s'agit pas d'utiliser le français parlé pour illustrer une théorie, mais de trouver une théorie qui permette d'aborder les données du français parlé. (Blanche-Benveniste & Jeanjean1987, 90)
>
> To me a corpus of any size signals a flashing neon sign "Think again", and I find it extremely difficult to fit corpus evidence into received receptacles.
> ... [T]he language obstinately refuses to divide itself into the categories prepared in advance for it ... (Sinclair 2001, 357-8)

References

AMIR N. SILBER-VAROD V. & IZRE'EL S. (2004), Characteristics of intonation unit boundaries in spontaneous spoken Hebrew: Perception and acoustic correlates, *in* Bel B. & Marlien I. (eds.), *Speech Prosody 2004, Nara, Japan, March 23-26, 2004: Proceedings*, SProSIG (ISCA Special Interest Group on Speech Prosody), 677-680.

ARIEL M. (2000), The development of person agreement markers: From pronouns to higher accessibility markers *in* Barlow M. and Kemmer K. (eds.), *Usage-based Models of Language*. Stanford, California, CSLI, 197-260.

AUSTIN J. L. (1962), *How To Do Things With Words*, The William James Lectures delivered at Harvard University in 1955, second edition Edited by J. O. Urmson and M. Sbisa, 1975, Cambridge, MA, Harvard University Press.

BECKMAN M. E. & PIERREHUMBERT J. B. (1986), Intonational structure in Japanese and English, *Phonology Yearbook 3*, 255-309.

BEL B. & MARLIEN I. (2002), *Speech Prosody 2002, Aix-en-Provence, France — 11-13 April 2002: Proceedings*, Aix-en-Provence, Laboratoire Parole et Langage - UMR 6057 CNRS.

BEL B. & MARLIEN I. (2004), *Speech Prosody 2004, Nara, Japan — March 23-26, 2004: Proceedings*, SProSIG (ISCA Special Interest Group on Speech Prosody).

BIBER D., JOHANSSON S., LEECH G., CONRAD S. & FINNEGAN E. (1999), *Longman Grammar of Spoken and Written English*, Harlow, Longman.

BLANCHE-BENVENISTE C. (1975), *Recherches en vue d'une théorie de la grammaire française: Essai d'application à la syntaxe des pronoms*, thèse présentée devant l'Université de Paris III, le 23 juin 1973, Paris, Librairie Honoré Champions.

BLANCHE-BENVENISTE C. (1997), *Approches de la Langue Parlée en Français*, Collection l'essentiel Français, Gap-Paris, Ophrys.

BLANCHE-BENVENISTE C. (2000), Transcription de l'oral et morphologie, *in* Guille M. & Kiesler R. (eds.), *Romania Una et Diversa: Philologische Studien für Theodor Berchem zum 65., Geburstag*, Band 1: Sprachwissenschaft, Tübingen, Gunter Narr, 61-74.

BLANCHE-BENVENISTE C. & JEANJEAN C. (1987), *Le Français Parlé: Transcription et Edition*, Paris, Érudition.

BLANCHE-BENVENISTE C., DELOFEU J., STÉFANINI J. & VAN DEN EYNDE K. (1984), *Pronom et Syntaxe: L'approche Pronominale et son Application à la Langue Française*, Paris, SELAF.

BLANCHE-BENVENISTE C., BILGER M., ROUGET C., VAN DEN EYNDE K. & MERTENS, P. (1990), *Le Français Parlé: Etudes Grammaticales*, Sciences du Langage, Paris, CNRS.

BRAZIL D. (1995), *A Grammar of Speech*, Describing English Language, Oxford, Oxford University Press.

BRAZIL D. (1997), *The Communicative Value of Intonation in English*, Cambridge, Cambridge University Press.

CAHANA-AMITAY D. & KATSENBERGER I. (1999), Intonation units as processing units in spoken discourse, *in* Aisenman R. A. (ed.), *Developing Literacy Across Genres, Modalities, and Language*, Tel Aviv, Tel Aviv University Press, 199-207.

CHAFE W. (1994), *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*, Chicago, University of Chicago Press.

CHENG W. GREAVES C. & WARREN M. (2005), The creation of prosodically transcribed intercultural corpus: The Hong Kong Corpus of Spoken English (prosodic), *International Computer Archive of Modern English (ICAME) Journal 29*, 5-26.

COFFIN E. & BOLOZKY S. (2004), *A Reference Grammar of Modern Hebrew*, Cambridge, Cambridge University Press.

COHEN (2004), *Veshe'eno Yode'a Lish'ol — Ma U Omer? Ways of Asking Questions in Spontaneous Spoken Hebrew*, MA thesis, Tel Aviv University. (Hebrew with English abstract)

CRESTI E. & MONEGLIA M. (2004), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*, Studies in Corpus Linguistics 15, Amsterdam, Bejnamins.

CRESTI E. (2004), Notes on lexical strategy, structural strategies and surface clause indexes in the C-ORAL-ROM spoken corpora, *in* CRESTI E. & MONEGLIA M. (2004), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*, Studies in Corpus Linguistics 15, Amsterdam, Bejnamins, 209-256.

CROFT W. (1995), Intonation units and grammatical structure, *Linguistics 33*, 839-882.

CROFT W. (2005), Intonation units and grammatical structure in Wardaman and English, http://lings.ln.man.ac.uk/Info/staff/WAC/Papers/WardamanIUs.pdf, September 30, 2005.

CROWDY S. (1994), Spoken corpus transcription, *Literary and Linguistic Computing 9*, 25-28.

CRUTTENDEN A. (1997), *Intonation*, Second edition, Cambridge Textbooks in Linguistics, Cambridge, Cambridge University Press.

CRYSTAL D. (1997), *The Cambridge Encyclopedia of Language*, Second edition, Cambridge, Cambridge University Press.

DAMOURETTE J. & PICHON E. (1911-1940), *Essai de Grammaire de la Langue Française*, 7 vols., deuxième édition, 1968, Paris, D'Artrey.

DU BOIS J. (1987), The Discourse Basis of Ergativity, *Language 63*, 805-855.

DU BOIS J. (1991), Transcription design principles for spoken discourse research, *Pragmatics 1*, 71-106.

DU BOIS J. W. (2004), *Representing Discourse*, ms, Linguistics Department, University of California, Santa Barbara.

DU BOIS J. W., CUMMING, S., SCHUETZE-COBURN, S. & PAOLINO D. (1992), *Discourse Transcription*, Santa Barbara Papers in Linguistics, 4, Santa Barbara, CA, Department of Linguistics, University of California, Santa Barbara.

DU BOIS J. W., SCHUETZE-COBURN, S., CUMMING, S. & PAOLINO D. (1993), Outline of discourse transcription. *in* Edwards J. A. & Lampert M. D. (eds.). *Talking Data: Transcription and Coding in Discourse Research*, Hillsdale, New Jersey, Lawrence Erlbaum Associates, 45-89.

GIVÓN T. (2001), *Syntax: An Introduction*, I-II, Amsterdam, Benjamins.

GOLDENBERG G. (1998), On verbal structure and the Hebrew verb, *in* Goldenberg G., *Studies in Semitic Linguistics: Selected Writings*, Jerusalem, Magnes, 138-147.

GUSSENHOVEN C. (2004), *The Phonology of Tone and Intonation*, Research Surveys in Linguistics, Cambridge, Cambridge University Press.

HALLIDAY M. A. K. (1989), *Spoken and Written Language*, second edition, Oxford, Oxford University Press.

HALLIDAY M. A. K. (2004), *An Introduction to Functional Grammar*, third edition revised by Christian M. I. M. Matthiessen, London, Arnold.

HELASVUO M.-L. (2001), Emerging syntax for interaction: Noun phrases and clauses as a syntactic resource for interaction, *in* Selting M. & Couper-Kuhlen E. (eds.), *Studies in Interactional Linguistics*, Studies in Discourse and Grammar 10, Amsterdam, Benjamins, 25-50.

HIRST D. & DI CRISTO A. (1998), *Intonation Systems: A Survey of Twenty Languages*, Cambridge, Cambridge University Press.

HOCKETT CH. F. (1958), *A Course in Modern Linguistics*, New York, Macmillan.

IWASAKI S. & TAO H. (1993), A comparative study of the structure of the intonation unit in English, Japanese, and Mandarin Chinese, *Annual Meeting of the Linguistics Society of America, Los Angeles, CA, Jan. 9. 1993.*

IZRE'EL S. (2002), The Corpus of Spoken Israeli Hebrew: Textual samples, *Leshonenu 64*, 289-314.

IZRE'EL S. (2005), Transcribing spoken Israeli Hebrew: Preliminary notes, *in* Diskin Ravid D. & Bat-Zeev Shyldkrot, H. (eds.), *Perspectives on Language and Language Development: Essays in Honor of Ruth A. Berman*, Dodrecht, Kluwer, 61-72.

IZRE'EL S. & COHEN E. (2004), *Literary Old Babylonian*, Languages of the World/Materials 81, München, LINCOM Europa.

IZRE'EL, S. HARY B. & RAHAV G. (2001), Designing CoSIH: The Corpus of Spoken Israeli Hebrew, *International Journal of Corpus Linguistics 6*, 171-197.

IZRE'EL S. & RAHAV G. (2004), The Corpus of Spoken Israeli Hebrew (CoSIH); phase I: the pilot study, *in* Oostdijk N., Kristoffersen G. & Sampson G. (eds.). *LREC 2004 Satellite Workshop; Fourth International Conference on Language Resources and Evaluation: Compiling and*

Processing Spoken Language Corpora (Lisbon, Portugal), Paris, ELRA - European Language Resources Association, 1-7. (Downloadable at http://www.tau.ac.il/humanities/semitic/lrec2004.rtf, September 30, 2005)

JESPERSEN O. (1924), *The Philosophy of Grammar*, London, George Allen & Unwin (repr. 1951).

KENNEDY G. (1998), *An Introduction to Corpus Linguistics*, Studies in Language and Linguistics, London, Longman.

KHAN G. (2003), The study of Modern Hebrew syntax, *in* Hary B. (ed.). *Corpus Linguistics and Modern Hebrew: Towards the Compilation of The Corpus of Spoken Israeli Hebrew (CoSIH),* Tel Aviv, Tel Aviv University, The Chaim Rosenberg School of Jewish Studies, 157-718.

KIBRIK A. E. (2001), Subject-oriented versus subjectless languages, *in* Haplesmath M., König, E., Oesterreicher W. & Raible W. (eds.), *Language Typology and Language Universals*, Berlin, de Gruyter, 1412-1423.

KUZAR R. (1989), *Message Structure of the Sentence in Israeli Hebrew*, PhD dissertation, Jerusalem, The Hebrew University. (Hebrew with English summary)

LADD D. R. (1986), Intonational phrasing: The case for recursive prosodic structure, *Phonology Yearbook 3*, 311-340.

LADD D. R. (1996), *Intonational Phonology*, Cambridge Studies in Linguistics 79, Cambridge, Cambridge University Press.

LAMBRECHT K. (1994), *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents,* Cambridge Studies in Linguistics 71, Cambridge, Cambridge University Press.

LAUFER A. (1987), *Intonation*. Jerusalem, the Hebrew University of Jerusalem, the Institute of Jewish Studies, Hebrew Language Department. (Hebrew)

LYONS J. (1968), *Introduction to Theoretical Linguistics*, Cambridge, The University Press.

MASCHLER Y. (2001), *veke'ilu haragláyim sh'xa nitka'ot bifním kaze* ('and like your feet get stuck inside like'): Hebrew *kaze* ('like'), *ke'ilu* ('like'), and the decline of Israeli *dugri* ('direct') speech. *Discourse Studies 3*, 295-326.

MATSUMOTO K. (2003), *Intonation Units in Japanese Conversation: Syntactic, Informational, and Functional Structures*, Studies in Language Companion Series 65, Amsterdam, Benjamins.

MATTHEWS (1981), *Syntax*, Cambridge Textbooks in Linguistics, Cambridge, Cambridge University Press.

METTOUCHI A. (forthcoming a), Le problème du sujet en berbère, in Donabédian A. (ed.), *Cahiers de Linguistique de l'INALCO,* Publications des LANGUES'O.

METTOUCHI A. (forthcoming b), Word order in conversational Taqbaylit Berber: Preposed and postposed subjects, *in* Voigt, R. (ed.), *From Beyond the Mediterranean, 7. Internationaler Semitohamitisten-Kongreß*, Semitica et Semitohamitica Berolinensia 5.

MILLER J. & WEINERT R. (1998), *Spontaneous Spoken Language: Syntax and Discourse*, Oxford, Oxford University Press.

MONEGLIA M. (2004), The C-ORAL-ROM resource, *in* CRESTI E. & MONEGLIA M. (2004), *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*, Studies in Corpus Linguistics 15, Amsterdam, Benjamins, 1-70.

NESPOR M. & VOGEL I. (1986), *Prosodic Phonology*, Studies in Generative Grammar 28, Dodrecht, Foris.

OCHS E. (1979), Transcription as Theory, *in* Ochs E. & Schieffelin B. B. (eds.), *Developmental Pragmatics*, New York, Academic Press, 43-72.

PIERREHUMBERT J. & HIRSCHBERG J. (1990), The meaning of intonational contours in the interpretation of discourse, *in* Cohen Ph., Morgan J. & Pollak M. E. (eds.), *Intentions in Communication*, Systems Development Foundation benchmark series, Cambridge, Massachusetts, MIT, 271-311.

POLOTSKY H. J. (1944), *Études de syntaxe copte*, Le Caire, Publications de la Société d'Archéologie Copte. (=*Collected Papers*, Jerusalem, Magnes, 1971, 102-207.)

POLOTSKY H. J. (1962), Nominalsatz und Cleft Sentence im Koptischen, *Orientalia 31*, 413-430. (=*Collected Papers*, Jerusalem, Magnes, 1971, 418-435.)

RODRIGUE-SCHWARZWALD O. & SOKOLOFF M. (1992), A Hebrew Dictionary of Linguistics and Philology, Even-Yehuda, Reches. (Hebrew; English and Hebrew entries)

ROSÉN H. B. (1977), *Contemporary Hebrew*, Trends in Linguistics: State-of-the-Art Reports 11, The Hague, Mouton.

ROSÉN H. B. (1966), *Ivrit Tova* ("Good Hebrew"), second edition, Jerusalem, Kiryat Sefer. (Hebrew)

RUBINSTEIN E. (1968), *The Nominal Sentence: A Study in the Syntax of Contemporary Hebrew,* Tel Aviv, Hakibbutz Hameuchad. (Hebrew)

SCHWARZWALD O. R. (2001), *Modern Hebrew*, Languages of the World/Materials 127, München, LINCOM Europa.

SELKIRK E. O. (1984), *Phonology and Syntax: The relation between Sound and Structure*. Cambridge, Massachusetts, MIT.

SHATIL N. (1994), *Sentence Patterns of Spoken Hebrew*, MA Thesis, Tel Aviv University. (Hebrew; English summary)

SHISHA-HALEVY A. (1986), *Coptic Grammatical Categories: Structural Studies in the Syntax of Shenoutean Sahidic*, Analecta Orientalia 53, Roma, Pontificium Institutum Biblicum.

SILBER-VAROD (2005), *Characteristics of Prosodic Unit Boundaries in Spontaneous Spoken Hebrew: Perceptual and Acoustic Analysis*, MA thesis, Tel-Aviv University. (Hebrew; English abstract)

SINCLAIR (2001), Review of Biber D., Johansson S., Leech G., Conrad S. & Finnegan E. (1999), *Longman Grammar of Spoken and Written English*, Harlow, Longman, *International Journal of Corpus Linguistics 6*, 339-359.

STEEDMAN M. (2001), *The Syntactic Process*, Language, Speech, and Communication, Cambridge, Massachusetts, MIT.

TAO H. (1996), *Units in Mandarin Conversation: Prosody, Discourse, and Grammar*, Studies in Discourse and Grammar 5, Amsterdam, Benjamins.

VAN VALIN R. D. Jr. & LAPOLLA R. J. (1997), *Syntax: Structure, Meaning and Function*, Cambridge Textbooks in Linguistics, Cambridge, Cambridge University Press.

WALDMAN N. M. (1989), *The Recent Study of Hebrew: A Survey of the Literature with Selected Bibliography*, Winona Lake, IN, Eisenbrauns.

WOUK F. (ms), The syntax of intonation units in Sasak, unpublished manuscript.

YATZIV-MALIBERT I. (2002), *Méthodologies pour la description de quelques phénomènes syntaxiques de langue parlée: application à l'hébreu moderne*, Thèse de doctorat, École Pratique des Hautes Études: Sciences historiques et philologiques, [Paris.]