The affixal nature of the definite article in Hebrew

Shuly Wintner
Seminar für Sprachwissenschaft
Universität Tübingen
Kl. Wilhelmstr. 113
72074 Tübingen, Germany
shuly@sfs.nphil.uni-tuebingen.de

Abstract

This paper provides evidence for the affixal nature of the Hebrew definite article, ha-. Usually considered a stand-alone word or a clitic, the rules that govern the combination of the article with nominals are viewed as part of the syntax. This paper suggests an alternative analysis by which it is attached to its host nominals through a morphological process. Several arguments, from all levels of linguistic representation, are provided. First, accepted criteria for wordhood reveal the clear dependent behavior of the definite article. Then, its affixal characteristics are discussed: phonological, morphological, syntactic and semantic considerations support the suggested view. Furthermore, it is shown that treatment of the definite article as an affix facilitates the construction of both syntactic and semantic theories for Hebrew. In the framework of Head-Driven Phrase Structure Grammar, a single lexical rule accounts for the combination of the article with various types of nominals. An HPSG analysis of Hebrew noun phrases is sketched, in which the affixal view of the definite article is integrated.

1 Introduction

This paper investigates some of the peculiar properties of the Hebrew definite article, 'ha-'.[1] Most analyses of Modern Hebrew (see references in section 4) consider ha-, implicitly or explicitly, to have approximately the same syntactic and semantic contribution as English 'the'. In particular, the construction of definite nominals is considered a syntactic process. In contrast to the traditional view, we claim in this paper that ha- is best regarded as an *affix* rather than as a full-fledged word or even as a clitic. In particular, its combination with nominals takes place in the lexicon, so it is inaccessible to syntactic processes.

We pursue in this paper a linear approach to grammar, along the lines of Di Sciullo and Williams (1987) and Anderson (1992): we assume that morphology is separate from syntax, and furthermore that morphological processes take place before syntactic ones. We thus presuppose a view according to which, if an element is a *word*, or a *clitic*, then its combination with other elements takes place in the *syntax* and yields *phrases*; if an element is an *affix*, its combination with other elements takes place in the *lexicon* and yields *words*. Clitics, after Anderson (1992), are syntactic words which lack the prosodic properties to be words at the prosodic level, and are consequently post-syntactically attached to adjacent words. We assume that words are the atomic elements for syntax, and that the operations that words can undergo take place in the lexicon. Then, after all the morphological processes have

[1] We use in this paper a transcription of Hebrew, known as *Phonemic Script* (Ornan, 1994), that was accepted as a standard (number ISO-DIS 259-3). The definite article attaches, in the Hebrew script, to the word immediately following it, but in the Phonemic Script it forms a separate unit. We are consistent with the standard in the examples below, but we intend to show that the article actually should be viewed as a part of the word it attaches to. In the glosses, gender, number and person are indicated by 'f' (feminine), 'm' (masculine), 'pl' (plural), 'sg' (singular) and '3rd' (third person).

taken place, syntactic rules combine words into phrases. In the case of the Hebrew definite article, we show that it should be lexically attached to its host, rather than be subjected to syntactic rules.

The analyses presented in this paper are conveyed in the context of Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1994). One of the main advantages of using HPSG is that the theory lends itself very naturally to computational implementation. Indeed, the analyses described herein were tested and their predictions verified. The use of HPSG provides means for elegant, concise analyses to be made.

Section 2 describes the data that are relevant for understanding the problem. In section 3 we show that an affix view of the definite article is consistent with the accepted criteria for wordhood. First, we show that ha- should not be considered a word, based on phonological,morphological,syntacticand semanticgrounds. We then applyseveral tests for distinguishing between clitics and affixes, showing that they support the suggested view. Finally, an affixal view of ha-, and in particular its implication, namely that the combination of the definite article with nominals is a lexical process, facilitates the construction of syntactic analyses for Hebrew. Section 4 sketches an HPSG analysis of a fragment of Hebrew, concentrating on noun phrases, in which this view is integrated. Viewing definiteness as a morphological process, we stipulate a single lexical rule that accounts for the combination of the article with various types of nominals. We then show how this analysis is integrated with the rest of the data presented in the paper.

2 The Hebrew definite article

2.1 The structure of Hebrew NPs

Hebrew nouns are specified for *gender, number* and *person*[2]. While Hebrew is a relatively free constituent order language, the order of the elements in an NP is sometimes fixed. In particular, quantifiers (including determiners, cardinal numbers and the definite article) are pre-head; all the other adjuncts and complements are post-head. The possible complements are listed below by their default order in a noun phrase.

Determiners such as koll (all/every), robb (most-of), kamma (some) etc.

Cardinal numbers such as loa (three).

Definite article

Nominal complement Hebrew allows an elaborate system of nominal compounds. Those headed by nouns are described in section 2.2.

Adjectives Hebrew adjectives are marked for number, gender and definiteness, on which they must agree with the head noun.

Ordinal numbers such as $eni (second) are likewise marked.

Demonstratives such as ze (this-m), zo (this-f), 'elle (these) or 'ellu (those).

Possessives including possessive pronouns such as $elli (mine) as well as phrases ($ell dan – Dan's).

Subcategorized complements of derived (deverbal) and 'picture'-like nouns.

Prepositional phrases

Relative clauses

An example involving some of these complements is:

[2]Only pronouns are specified for person, other nouns are inherently third person.

(1) koll $e$$ ha- smalot ha- yapot ha- 'elle $elli mi- 'rhb
all six the dresses the nice the these mine from USA
'all these six nice dresses of mine from USA'

Relevant to the case in hand is the fact that many, though not all, of the modifiers, can be explicitly or implicitly definite. Hebrew marks *definiteness* in a way that differs much from, say, English or German (but resembles other Semitic languages, notably Arabic, and also some Balkan and Scandinavian languages, as will be noted below). There is only one definite article in Hebrew, ha-, it does not inflect and it attaches (pre-nominally) to *words*, not to phrases. It can combine with various kinds of nominals: common nouns, proper nouns, adjectives, ordinal numbers, cardinal numbers and demonstratives. Moreover, definite noun phrases in Hebrew are *polydefinite*: most of the elements of the phrase are required to be explicitly definite, and there is a strict requirement that these elements *agree* on definiteness for the phrase to be grammatical. Hebrew arguably has *indefinite* articles[3] ('exxad, 'axxat, 'xadim), but their use is optional and not common. It is therefore useful to view bare nominals (with no attached definite article) by default as indefinite. The data are summarized[4] in (2). The basic (in)definite noun phrase in Hebrew is demonstrated in (2a). In (2b) the noun is modified by an adjective, and agreement on definiteness is demonstrated. The same pattern recurs for ordinals (2c) and demonstratives (2d).

(2) **a.** sepr ('exxad) / ha- sepr
book (one) / the book
'a book / the book'

b. sepr gadol / ha- sepr ha- gadol / *sepr ha- gadol / *ha- sepr gadol
book big / the book the big / book the big / the book big
'a big book / the big book'

c. sepr $eni / ha- sepr ha- $eni / *sepr ha- $eni / *ha- sepr $eni
book second / the book the second / book the second / the book second
'a second book / the second book'

d. sepr ze / ha- sepr ha- ze / *sepr ha- ze / *ha- sepr ze
book this / the book the this / book the this / the book this
'this book / this book'

2.2 Noun–noun constructs

Hebrew nominals come in two forms: the *absolute* form is used in any context; the other form, known as *construct* or '*nismak*' form, is used only in the context of noun–noun constructs. For many nominals, especially among singular masculine and plural feminine, the two forms are identical; for many others the construct form is phonologically reduced.[5] Construct forms exist for all common nouns and adjectives and for cardinals up to ten. Some examples follow:

(3)

absolute:	sepr	sparim	xulca	xulcot	loa	alo	gadol	gdola
construct:	sepr	siprei	xulcat	xulcot	lot	lo	gdol	gdolat
	book	books	shirt	shirts	three-m	three-f	big-m	big-f

This section covers noun constructs only; adjectival constructs are discussed in section 2.3.

A noun–noun construct is a phrase consisting of a construct-state noun followed by a noun phrase. We refer to the first element of this construction as the *head* and to the second – as the *complement*.

[3] This analysis is due to Ornan (1964) and is challenged by Rosén (1977, p. 155). For a detailed discussion see Givón (1981).

[4] Some phrases, such as ha- sepr gadol or ha- sepr $eni, are grammatical *sentences* but not noun phrases. We mark such phrases as ungrammatical below to indicate that they are unacceptable as noun phrases.

[5] While the morphological rules relating absolute and construct forms are too complex to discuss here, they are relatively easy to formalize.

The phrase inherits all the morpho-syntactic features of the head, with the exception of *definiteness*, which is inherited from the NP complement (Borer, 1984, pp. 41-68). Semantically, the relation between the head and the complement is usually that of possessed-possessor, but various other relations are possible (see Levi (1976) for a detailed survey). Consider the following examples (c denotes the construct-form):

(4) **a.** pirxei gann yapim parxu
flowers-pl-c garden-sg beautiful-pl flourished-pl-past
'beautiful garden flowers flourished'

b. pirxei ha- gann ha- yapim parxu
flowers-pl-c the garden-sg the beautiful-pl flourished-pl-past
'the beautiful garden flowers flourished'

The properties of the NPs in both of these cases can be determined by the verb, since the subject and the main verb must agree on number and gender in Hebrew. Thus it is clear that the head of the NP in both cases is pirxei rather than gann. The fact that the adjective yapim is in plural indicates that it modifies the head pirxei, rather than gann, as adjectives must agree with the head they modify on number, gender and definiteness. However, this head is not definite in any of the examples; rather, it is the complement ha- gann that is definite in (4b), but indefinite in (4a), in agreement with the definiteness of the adjective. Hence definiteness is inherited from the complement, rather than the head.

Another technique for checking the definiteness of an NP in Hebrew involves the direct object (accusative) marker, 'et, that is glossed as 'ACC'. Having the characteristics of a preposition, 'et introduces only definite NPs. Hence:

(5) qaniti pirxei gann / qaniti 'et pirxei ha- gann
I-bought flowers-c garden / I-bought ACC flowers-c the garden
'I bought garden flowers / I bought the garden flowers'

*qaniti 'et pirxei gann
I-bought ACC flowers-c garden

The process of compounding is recursive, as the resulting phrase is a legitimate NP for combining with some other construct form. When more than two nouns are combined, the resulting phrase's definiteness is determined by the last (rightmost) NP, the one in absolute form, while the head features are percolated from the first (leftmost) noun, which is of course in construct:

(6) yaldei mnahhel taxnot ha- rakkebt roqdim
children-m-pl-c manager-m-sg-c stations-f-pl-c the train-f-sg are dancing-m-pl
'the train stations manager's children are dancing'

It is important to note that nominals in construct-form can *only* occur in the context of N-N constructs: a construct-noun with no immediate NP succeeding it is ungrammatical:

(7) ba- rxob holekt yalda / *yaldat
in-the street walks girl / girl-c
'a girl walks in the street'

2.3 Adjective–noun constructs

Noun–noun constructs, discussed in the previous section, have received much consideration in the literature. However, a similar phenomenon, namely adjective-headed constructs, is much less known (but see a brief discussion in Hazout (1991, pp. 123-130)). In such phrases, the adjective must be in the construct form, and it must be immediately complemented by a noun. The resulting phrase is an adjectival phrase (ADJP). Consider the data in (8):

(8) **a.** yalda gdolat &einaym / ha- yalda gdolat ha- &einaym
girl-f-sg big-f-sg-*c* eyes-pl / the girl big-f-sg-*c* the eyes-pl
'a big-eyed girl / the big-eyed girl'

b. yladot gdolot/∗gdolat/∗gdolei &einaym
girls-f-pl big-f-pl-*c*/big-f-sg-*c*/big-m-pl-*c* eyes-pl
'big-eyed girls'

c. ∗ha- yalda gdolat &einaym / ∗yalda gdolat ha- &einaym
the girl big-f-sg-*c* eyes-pl / girl big-f-sg-*c* the eyes-pl

These data suggest an analysis that is very much similar to the one accounting for noun–noun constructs. In particular, they show that the head features of the phrase are inherited from the first (leftmost) element: first and foremost, the category is that of the first element, i.e., an ADJP. The agreement features (gender and number) of the phrase are inherited from the head, as shown in (8b); however, the definiteness is inherited from the nominal complement, as shown in (8c).

In spite of the above, there still is one important, rather puzzling difference between nouns and adjectives in construct form. While the former can combine with *any* NP, the latter require *words* as their complements. The data in (9) illustrate this point.

(9) **a.** yaldei ha- &olam
children-*c* the world
'the world's children'

b. gdolat ha- &einaym
big-f-*c* the eyes
'the big eyed'

a. yaldei ha- &olam ha- gadol/lii/ze
children-*c* the world the big/third/this
'[the big world / the third world / this world]'s children'

b. ∗gdolat ha- &einaym ha- yruqqot/$niyot/'elle
big-f-*c* the eyes the green-pl/second/these

a. yaldei koll ha- &olam
children-*c* all the world
'[all the world]'s children'

b. ∗gdolat koll ha- &einaym
big-f-*c* all the eyes

3 What's in a word?

While an agreed upon definition for the terms *affix*, *clitic* or *word* is still unavailable, the following quotations from Spencer (1991) seem to be uncontroversial:

> [p. 21] "Inflectional operations leave untouched the syntactic category of the base, but they add... meaning... and also grammatical function... The two most widespread and important types of grammatical function served by inflection are agreement and government."

> [p. 350] "Clitics are elements which share certain properties of fully fledged words, but which lack the independence usually associated with words. In particular, they can't stand alone, but have to be attached phonologically to a *host*... Typically, clitics are function words... They are generally assumed to be incapable of bearing stress or accent."

The main claim of this paper is that the Hebrew definite article should be viewed as an affix. We start with a set of criteria that are accepted tests for distinguishing words from bound morphemes. Once the dependent status of the definite article in Hebrew is established, we discuss its affixal properties in section 3.3.

In a very influential paper, Zwicky (1977) lists six criteria for distinguishing between affixes and words. The criteria, which are pre-theoretical in nature, are elaborated upon in Zwicky (1985), where several tests for distinguishing clitics from independent words are listed. The tests are suggested to be taken as "symptoms" of a linguistic state of affairs, rather than as necessary and sufficient conditions for it. Still, when a certain element complies with many of the tests, it is most likely to be in this particular state of affairs. As we show below, ha- passes most of the tests for clitic-hood, and – just as importantly – fails hardly any of them (some are inapplicable to the case in study). We list the criteria below, applying each of them to the Hebrew definite article (some of the criteria which are irrelevant to this case ore omitted).

3.1 Zwicky's tests

Sandhi: Internal sandhi rules apply only within words, external sandhi rules apply only between phonological words and not within them.

The definite article, which is pronounced [ha], changes to [he] when combined with certain nominals (roughly, those whose first syllable is 'ha' or '&a'). Thus, /ha- harrim/ (the mountains) is pronounced [heharim]. This is a word-internal rule: in general, guttural sounds fail to geminate in Hebrew; as a compensation, in contexts that require gemination of a guttural sound, the preceding vowel is changed. If one assumes that attaching ha- to a word in Hebrew requires gemination of the initial sound of this word, the two phenomena described above are immediately obtained. Of course, such an assumption can only be made if ha- is taken to be an affix.

Accent: Clitics are accentually dependent, that is, an element which does not bear an accent of its own is probably a clitic.

Phonologically, ha- does not bear an independent stress.

Binding: Bound elements are clitics; if an element is bound, and especially if it cannot occur in complete isolation, it should be a clitic.

This is indeed the case with ha-. Never, under no circumstances, can it stand in isolation.

Construction: If the distribution of an element is correctly stated in terms of its ability to combine with single words, it will be a clitic. Certainly, if an item combined syntactically only with single words, we should hesitate to classify it as a word.

Note that the construction criterion is 'one-way': it can only support a clitic analysis. As is clear from the examples in (2), the definite article usually combines with single words.

Distribution: Clear cases of clitics typically have distributions describable by single principles.

As we show below, it is easy to stipulate a single principle for the distribution of ha-: it combines with nominals that are inherently indefinite. The only exception is that the rules governing its combination with proper names are complex.

Complexity: Words frequently are morphologically complex; clitics rarely are.

Clearly, ha- behaves like a clitic in this respect.

Deletion: Proper parts of words are not subject to deletion under identity; whole words may undergo such deletions.

The natural environment to apply this test is that of coordination; we elaborate on this issue in section 3.3.

Movement: Proper parts of words are not subject to movement rules: they cannot serve as gaps in gap-filler relations.

This complies well with the case of definite nominals: whenever the nominal moves, it moves with the attached article.

To these tests we add the peculiarity of the adjunct–noun construct phenomena, listed in (8) (section 2.3). What these data demonstrate is that the complement of a construct state adjective can be a noun, but not a noun phrase: this noun cannot be modified by adjectives, quantifiers, ordinals or demonstratives. It is interesting to note that regardless of whether 'exxad is considered to be an indefinite article or just a cardinal number, it cannot modify the complement of a construct-state adjective:

(10) ra'iti xatul yroqq &einaym / *raiti xatul yroqq &ayn 'axxat
I-saw cat green-*c* eyes / I-saw cat green-*c* eye one/a
'I saw a green-eyed cat / (putatively) I saw a cat with one green eye'

However, the modifier noun *can* be preceded by the definite article, as can be seen from the examples in (8). This observation supports our claim that definite nouns are not phrases in Hebrew.

3.2 The semantics of ha-

The usual semantic contribution of definiteness is determination: a definite noun phrase denotes a unique entity. *Compositional* approaches to semantics combine the meaning of the definite article with that of the NP to yield a unique reading of the definite NP. If ha- were a word, this should have been the case for Hebrew, too. However, in Modern Hebrew *definiteness* and *determination* are not parallel: there are many contexts in which definite and indefinite NPs have identical meanings; there are also cases of determination that are not carried out through the use of the definite article. Any attempt to combine, compositionally, the semantic effect of ha- with the meaning of an indefinite nominal would lead to severe difficulties.

To demonstrate the differences between definiteness and determination, consider the following examples:

Demonstratives Hebrew nouns can be modified by demonstratives such as ze (this-m), zo (this-f), 'elle (these) or 'ellu (those). Semantically, such a modification results in a determination of the entity denoted by the noun. Syntactically, ha- *can* be added to the noun, in which case it must modify the demonstrative as well. However, it is not obligatory for this process to be encoded by the incorporation of the definite article. In other words, ha- is used in this context solely as an agreement marker, with no semantic contribution of its own:

(11) sepr ze nimkar hei@eb / ha- sepr ha- ze nimkar hei@eb
book this is-sold well / the book the this is-sold well
'this book sells well'

Generic nouns In many contexts, when denoting abstract entities, Hebrew nouns can be both definite and indefinite, with no change of meaning:

(12) (ha-) &i$$un mazziq la-bri'ut
(the) smoking harms to-the-health
'smoking is hazardous for the health'

(ha-) $oxd y&awwer &einei xkamim
(the) bribe will-blind eyes-*c* wise
'a bribe blinds the eyes of the wise'

Ordinals When ordinals are used to modify nouns that are already independently determined, they may or may not be preceded by the definite article:

(13) ra'iti $nei xtulim, (ha-) 'exxad $axor, (ha-) $eni laban
I-saw two cats (the) one black (the) second white
'I saw two cats, one black, the other white'

These examples indicate that the definite article in Hebrew does not always have a determining function; rather, it serves in several contexts simply as an agreement marker. Determination can be achieved by alternative mechanisms, not necessarily related to definiteness. Any compositional semantics approach to definite NPs in Hebrew would have to associate some semantic function with the definite article, if it is to be taken as a word, and would thus face severe difficulties in explaining the above data. It seems that a better approach would be to consider the definite article as marking the nominals it attaches to simply as 'definite'. Definiteness would then be an agreement feature, such as 'number', carrying no specific meaning.

It is beyond the scope of this paper to suggest a complete semantic analysis for definiteness in Hebrew. We simply indicate that since there is no simple one-to-one relation between determination and definiteness, and since clearly definite NPs can in some contexts denote exactly what their indefinite counterparts do, the definite article does not necessarily carry a semantic contribution of its own. Therefore, regarding a definite noun as a phrase would be problematic for any attempt to provide compositional meaning to phrases.

3.3 Clitics vs. affixes

Once the non-word status of the Hebrew definite article is established, we show that it is actually an affix, rather than a clitic. We start by applying to it the criteria listed in Zwicky and Pullum (1983), distinguishing clitics from inflectional affixes. Again, the criteria are stated in terms of "tendencies":

- Clitics exhibit low selectivity – affixes are more selective.

  As shown above, ha- can combine with all kinds of nominals, including nouns, adjectives, numerals and demonstratives. While this might seem a rather low degree of selectivity, note that these elements do form a natural class. For example, all of them can be used (elliptically) to denote an entity. Furthermore, note that ha- *never* attaches to, say, prepositions or adverbs. It also does not combine with quantifiers, which occur pre-nominally in Hebrew. Thus, we might say that the article is quite selective in the hosts it attaches to.

- Arbitrary gaps in the set of combinations are more characteristic of affixed words than of clitic groups.

  As far as we can see, there are no such gaps in the case of ha-: it combines with *all*[6] nouns and adjectives, for example.

- Morpho-phonological idiosyncrasies are more characteristic of affixed words than of clitic groups.

  As shown above, such cases are present in the combination of ha- with nominals; they usually affect the article, but rarely also the nominal.

- Semantic idiosyncrasies are more characteristic of affixed words than of clitic groups.

  As section 3.2 reveals, such cases occur with definite nominals.

As can be seen, these criteria do not unanimously suggest either view. However, this is often the case; in particular, as Miller (1992) notes, due to the possibility of regular inflectional systems, the phonologically and morphologically oriented criteria can only provide arguments in favour of a lexically attached affixal status, not against it. That is, the fact that there are no arbitrary gaps, or no morpho-phonological idiosyncrasies, does not imply anything. To allow a more direct argumentation in favour of a clitic status, Miller (1992) suggests the following criterion: if an item must be repeated on each conjunct in a coordinate structure, then it must be an affix and cannot be a clitic; if it cannot

[6] The definite article does not combine with construct state nominals, but this is not an *arbitrary* gap, as we show in section 4.3.

be repeated, it must be a clitic and cannot be an affix. If repetition is optional, no evidence can be drawn.

This test is easy to apply to the case of ha-. First, note that coordination of elements to which ha- attaches is possible in Hebrew:

(14) qaniti sepr w- maxbert
I-bought book and notebook
'I bought a book and a notebook'

&einaym gdolot w- yruqqot
eyes big and green
'big green eyes'

When the elements are definite, ha- cannot have wide scope over the coordination, but rather must be repeated for each of the conjuncts:

(15) qaniti 'et ha- sepr w- ha- maxbert
I-bought ACC the book and the notebook
'I bought the book and the notebook'

ha- &einaym ha- gdolot w- ha- yruqqot
the eyes the big and the green
'the big green eyes'

An omission of one of the occurrences of ha- results either in ungrammaticality or in a different reading, in which the article has a narrower scope:

(16) qaniti 'et ha- sepr w- maxbert
I-bought ACC the book and notebook
'I bought the book and a notebook'

*ha- &einaym ha- gdolot w- yruqqot
the eyes the big and green

Miller (1993) suggests three more criteria. Assuming the lexicalist hypothesis, namely that affixes are lexically attached to their stems, it immediately follows that if lexical phonological rules can apply to the unit formed by the element in question and the host it attaches to, then the element is an affix. In other words, if the realization of the element depends on the morphological structure of the host, it must be an affix. We have shown such dependencies in section 3.1. The second criterion has to do with processual realization of the morphological element and hence is inapplicable to the case of ha-. The third criterion deals with haplology: if only one occurrence of the element is present where one would expect two (because of syntactic structure), then the element is an affix. We could not find any such evidence in the case of definite nominals in Hebrew, so that this criterion does not contribute to either view.

The coordination criterion is one among many applied by Börjars (1992) to determine the status of morphological definiteness markings (DEFs) in Scandinavian and Balkan languages. Some of the criteria draw a clear line between these two groups of languages: DEF is an affix in the Scandinavian languages (Swedish, Norwegian, Danish, Faroese and Icelandic), a clitic in the languages of the Balkan (Albanian, Romanian, Macedonian and Bulgarian). However, as Börjars (1992) notes, the coordination criterion seems to have different predictions: DEF is not duplicated on conjuncts in any of the languages investigated. While this might have reduced the relevance of the coordination test, it might also serve as an indication that DEF in the Balkan languages is actually an affix. Indeed, this claim exactly is made by Halpern (1992): comparing the definite article of the Balkan languages with other

second-position clitics, it is shown that an affix view of DEF in these languages is favourable. Thus, the relevance of the coordination criterion is retained.

To summarize, clitics are entities that show some of the characteristics of words, and some of affixes. Therefore there are different kinds of clitics, covering the spectrum between words and affixes. It should be clear from the tests applied in this section that the definite article in Hebrew is much closer to an inflectional affix than to a stand-alone word. In any case, it combines with its hosts as a result of a lexical, not a syntactic, process.

4 An analysis of definiteness in Hebrew

The purpose of this section is to show that an affix view of the definite article facilitates simpler, more natural accounts for the structure of noun phrases in Hebrew. Existing analyses of the definite article in Hebrew, conveyed in Chomskian frameworks, tend to view it as a head, and its combination with nominals as a syntactic process. Ornan (1964) takes the article to be a full-fledged word. Ritter (1988) views it as a clitic; the account of definiteness is different in Ritter (1991), but the status of the definite article is not explicitly changed. Siloni (1991) criticizes the clitic analysis and suggests one by which the noun incorporates with the article after being moved. Others (Shlonsky, 1990; Siloni, 1994) take the article to be a major participant in syntactic processes. Space limitations prevent us from demonstrating the problems involved in such a view (see Borer (1994), Wintner (1998b) for more details). Many of the arguments in favour of the DP hypothesis and its application to Hebrew are theory internal; the analysis we suggest is carried out in the framework of HPSG (Pollard and Sag, 1994) in which movements are ruled out and empty (phonologically null) categories are discouraged, and we claim that an NP analysis for noun phrases is more natural. The fuller data presented in this paper, combined with the different theoretical framework, yield a much simpler analysis.

4.1 The framework

HPSG is formulated as a set of constraints on *typed feature structures*; these are used to represent *signs* (both words and phrases). For example, a *noun* is a word whose HEAD feature has the type *noun*. Figure 1 depicts the lexical entry[7] of the common noun perx (flower), where '⟨...⟩' denotes a list. The feature DEP encodes prosodic dependency (Wintner, 1998a; Wintner, 1998b); it indicates an expectation for an *immediate* complement that cannot be extracted. We treat possessors as complements, following (Borsley, 1989; Borsley, 1995), and hence the COMPS list of this noun contains one element, a genitive PP (see Wintner (1998a) for more details). The feature DEF is discussed in section 4.2.

$$\left[\begin{array}{ll} \textbf{word} & \\ phon: & \text{perx} \\ synsem: & \left[\begin{array}{ll} \textbf{synsem} & \\ cat: & \left[\begin{array}{ll} \textbf{cat} & \\ head: & \left[\begin{array}{l}\textbf{noun}\\ def: -\end{array}\right] \\ subj: & \langle\rangle \\ comps: & \langle \text{PP[of]} \rangle \\ spr: & \langle\rangle \end{array}\right] \end{array}\right] \\ dep: & - \end{array}\right]$$

Figure 1: The lexical entry of the noun perx (flower)

HPSG "rules" are organized as a set of *principles* that set constraints on the properties of well-formed phrases, along with a set of *ID schemata* that license certain phrase structures. The schemata

[7] As the semantics of definiteness is not addressed in this paper, the values of the CONTent feature are systematically suppressed in the depiction of feature structures. Also, we sometimes contract the path SYNSEM|LOCAL|CAT to SYNSEM|CAT or just CAT.

are independent of the categories of the involved phrases; they state general conditions for the construction of larger phrases out of smaller ones, according to the function of the sub-phrases. Suppose the Hebrew definite article were a *word*, or in other words, that its combination with a nominal were a syntactic process. What would the syntactic relation between ha- and the nominal be? Which of the schemata would apply for this combination?

The standard analysis for English, presented in Pollard and Sag (1994, section 9.4), views articles as subcategorized complements of nouns. The article combines with the noun through the *specifier-head* schema: the noun is the head of the construction, and since HPSG requires that phrases be saturated, that is, have empty subject, specifier and complement slots, a bare noun (with no article) is rendered ungrammatical. While this might be appropriate for English, it certainly isn't for Hebrew – as the data in (2) above show, bare nouns function perfectly fine as complete NPs.

Following the work of Abney (1987), the standard HPSG analysis for German now considers article-noun combinations to be DPs, rather than NPs (Netter, 1994). Preferring any of the two analyses, in the context of HPSG, boils down to deciding whether it is the article or the noun that is the *head daughter* of a nominal phrase. In Wintner (1998b) we show in detail why the definite article cannot head a noun phrase.

A completely different approach is taken by Kolliakou (1996), accounting for definiteness in Modern Greek in the framework of HPSG. Similarly to Modern Hebrew, Greek has a system of polydefinites, but *monadic* definites are allowed, too. The definite article is viewed as an adjunct, that selects the nominals it attaches to through a MOD feature in its lexical entry, and marks them as definite. There's a minor problem with this approach: when the article is combined with an adjective, it marks the adjective as definite – but not the value of the MOD feature of the adjective. As a result, definite adjectives select indefinite heads. To overcome this problem, Kolliakou (1996) suggests that the definiteness marker of adjectives should be left unspecified in the lexicon, and an ad-hoc *uniqueness principle* would take care of definiteness agreement in every instance of a head-adjunct structure in which the adjunct is an adjective.

Since Hebrew depicts *only* phenomena of polydefinites, a more suitable account – along the same lines – would have been to treat ha- as a *marker*, rather than an adjective, and have it select indefinite nominals and mark them as definite. This solution is problematic in two respects: HPSG does not account for agreement processes in the grammar; all other agreement constraints are listed in the lexicon. Moreover, the uniqueness principle applies only to one certain ID schema, namely the head-adjunct one; we believe that principles must be as general as possible. A better generalization for the Hebrew data can be obtained in a different way.

To summarize, any attempt to account for the combination of ha- with a nominal in Hebrew using any of HPSGs phrase structure schemata is either inadequate or results in missing important generalizations: it necessitates stipulation of ad-hoc rules and constraints. On the other hand, if ha- is viewed as an affix rather than a word, its combination with nominals should be accounted for in the lexicon, possibly through a lexical rule such as the one we presently suggest. None of the principles has to be modified.

4.2 Definiteness as a lexical process

We seek in this section a generalization for the definiteness phenomena described in section 2 above. In particular, an immediate question presents itself: of all the elements in a Hebrew NP, which are the ones that can be definite, or in other words, which are the elements that must agree on definiteness with the head noun? Surely, the criteria cannot be semantic: if they were, it would have been impossible to explain why ha- can attach to the (inherently determined) demonstratives. A related question, one that has received much consideration in recent linguistic research, has to do with construct state nominals; namely, why can't the head of a construct be directly rendered definite?

The answer has to do with the nature of the definite article, and in particular with the following properties:

- ha- attaches to words, not to phrases;

- it attaches only to nominals, and to all kinds of nominals;
- it only combines with indefinite words.

It is crucial for this analysis that the process of adding the definite article take place in the lexicon; in particular, it takes place *before* other cliticization processes, such as personal pronoun cliticization, apply.

By *nominals* we mean nouns (e.g., sepr), adjectives (e.g., gadol), ordinals (e.g., $eni), cardinals (e.g., loa) and demonstratives (e.g., ze). A preliminary segment of the type hierarchy that captures this definition is given in figure 2.

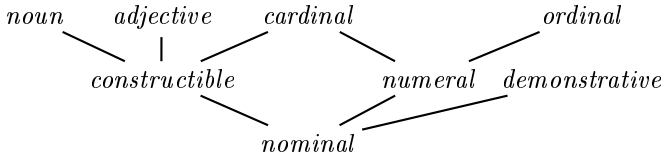

Figure 2: The *nominal* sub-hierarchy

An additional (boolean) feature, DEFiniteness, is required for encoding the value of definiteness in nominals. As definiteness agreement in Hebrew is not a semantic process, we add this feature to the CATegory of nominals (rather than to their CONTent). Since definiteness is a feature of phrases, inherited from the lexical head and participating in agreement relations, DEF is a head feature, appropriate for all *nominals*. Viewing definiteness as a lexical process, we introduce the *Definite Lexical Rule* (DLR). It operates on all nominal words, provided that the value of their DEFiniteness feature is '−'. In all categories its effect on the phonology is determined by the same phonological rules; we use the function *definite* to abstract over them. The DLR changes the value of the path SYNSEM|LOC|CAT|HEAD|DEF from '−' to '+'. *Adjuncts* specify the heads they select as the value of the MOD feature in their lexical entries. Like any other nominal, they have a DEFiniteness feature, whose value is shared with the value of the path MOD|LOC|CAT|HEAD|DEF. When the DLR operates on adjuncts, it results in a specification of a '+' value for both paths. Thus it is guaranteed that definite adjectives, for example, are not only specified as definite but also select definite heads. The DLR[8] is depicted in figure 3; its effect when applied to a few nominals is exemplified in figures 4 and 5. A discussion of construct-state nominals and their definiteness specification is deferred to section 4.3.

$$\left[\begin{array}{l}\textbf{word}\\ phon: \boxed{1}\\ synsem:cat: \left[\begin{array}{l}\textbf{cat}\\ head: \left[\begin{array}{l}\textbf{nominal}\\ def: -\end{array}\right]\end{array}\right]\end{array}\right] \mapsto \left[\begin{array}{ll} phon: & definite(\boxed{1})\\ synsem:cat:head:def: & +\end{array}\right]$$

Figure 3: The Definite Lexical Rule

Once the process of adding the definite article is taking place in the lexicon, the head-adjunct schema can remain intact (that is, no additional principles such as the uniqueness principle are needed). Moreover, the agreement in definiteness between a nominal and its adjuncts is stated in the lexical entry of the adjuncts, just like agreement on number and gender is.

[8] Two assumptions are implicit in this rule: (1) that it is triggered by subsumption, not unification (see Hinrichs and Nakazawa (1996) for a discussion and other examples of this concept); (2) that when a value at the end of a path is modified, all paths that are reentrant with it are modified as well.

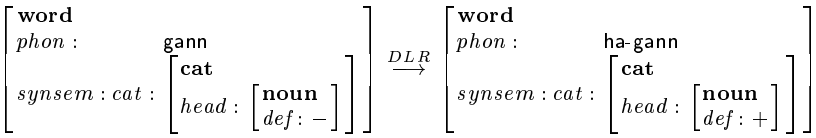

Figure 4: The effect of the Definite Lexical Rule on nouns

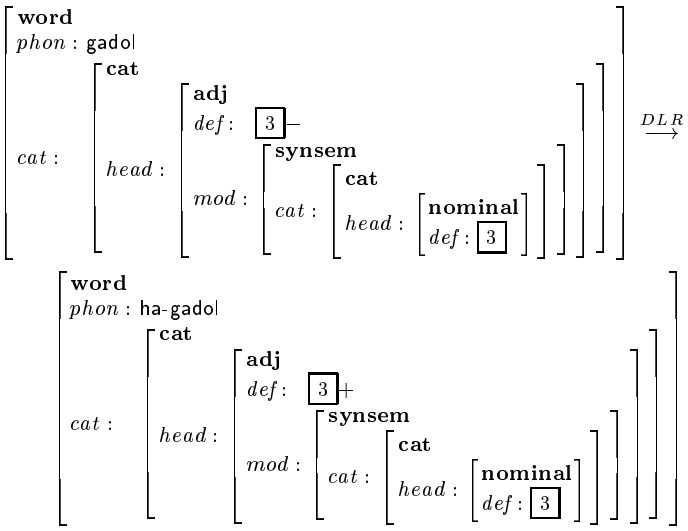

Figure 5: The effect of the Definite Lexical Rule on adjectives

4.3 Constructs

It is common practice in HPSG to account for so called "movement" phenomena by means of value sharing (reentrancies). This very solution is applicable to the case of construct state nominals in Hebrew, too. To explain the fact that such nominals cannot be rendered definite explicitly, but rather "inherit" the definiteness feature of their complements, Borer (1988), Ritter (1988), Shlonsky (1990) and Siloni (1994) all resort to an analysis by which the head noun must be raised from its base position. This results in awkward structures that are independently unmotivated. However, this phenomenon can be easily explained in a theory such as HPSG: construct state nominals are words, and their lexical entries must express an expectation for an immediate complement; that is, an indication (the SYNSEM value) of the compulsory complement of construct nominals is present in the lexical entry of the nominal. It is thus possible to share, in the lexicon, the values of the definiteness feature in both the nominal and its complement.

The construct form is generated from the absolute form by means of a morphological process. Apart from modifying the phonology[9] of the nominal, this process has a double effect. Recall that nouns are specified for a possessor in their COMPlements list; therefore, there is no need to add a subcategorized complement for construct state nouns. The rule only has to pick a complement from this list, and unify the values of the DEF feature of the nominal and the complement. In addition, the rule marks the nominal as 'DEP +', to indicate the fact that construct state nominals are prosodically dependent. When the nominal is combined with its complement, the resulting phrase inherits the

[9] The function *phon_reduce* computes the phonology of the construct noun.

definiteness from the latter. Figure 6 depicts the effect of this process when applied to nouns. Notice that the results of this process, i.e., the lexical entries of construct-state nouns, are not specified as 'DEF −' (in fact, they are not specified for definiteness at all). Consequently, the definiteness lexical rule stated above cannot apply to them (recall the requirement that the entry be subsumed by the antecedent of the rule). The fact that construct state nominals cannot be rendered definite directly is naturally obtained.

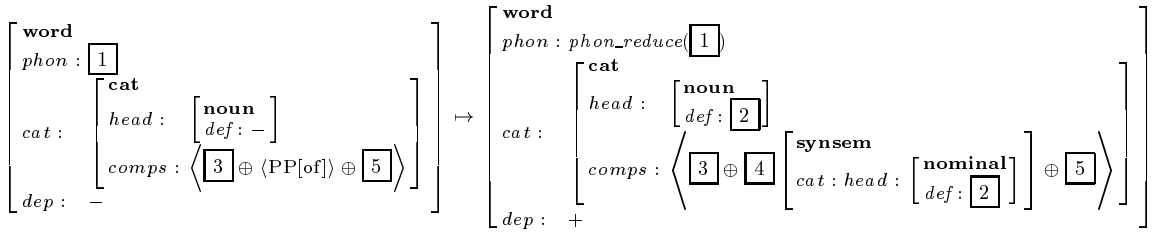

Figure 6: The relation between absolute and construct forms

Once this process is applied to construct-form nouns, their lexical entry specifies that they expect a nominal complement. Noun–noun constructs can thus be constructed by the head-complement schema. Since the DEF value of the construct head and its complement are shared, and since DEF is a head feature, it is also shared by the mother; thus, the DEF feature of the phrase is inherited from the complement, as required. Figure 7 depicts this process; notice in particular how the definiteness of the phrase is inherited from the complement using a reentrancy in the head.

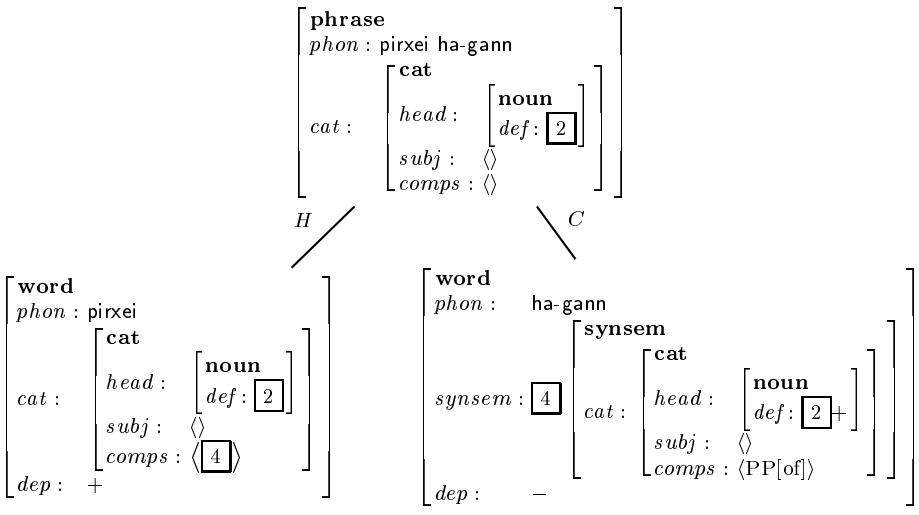

Figure 7: A construct-state NP

The striking similarities between noun–noun and adjective–noun constructs imply that they are actually only two instances of one process: any analysis that would suggest two different mechanisms to account for both phenomena is bound to be redundant. We simply extend the analysis of noun–noun constructs delineated above to the case of construct-state adjectives: such adjectives are lexically specified to subcategorize for nouns. They cannot occur independently, with no immediate complement, and hence are marked as dependent; the phrase is constructed through the head-complement schema. To account for the only difference between nouns and adjectives in this respect, the noun on

the COMPS list of construct-adjectives is required to be a *word*.[10] Figure 8 depicts a derivation of the adjective–noun construct gdolat &einaym (big eyed).

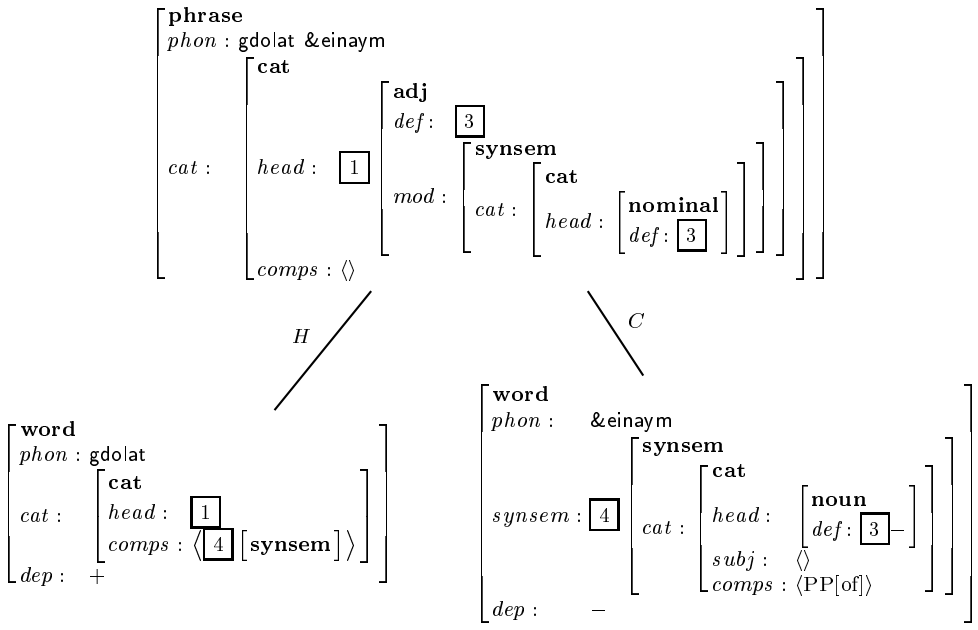

Figure 8: A construct-state ADJP

5 Conclusions

The main claim of this paper is that the definite article in Hebrew is better viewed as an affix than as a stand-alone word or even a clitic. We have applied several accepted tests for wordhood, from all levels of linguistic representation (phonology, morphology, syntax and semantics) to the case of the definite article and showed that the vast majority of them suggest the proposed view. We then showed the advantages of our analysis for both syntactic and semantic theories of Hebrew: if the definite article is considered a word, a certain syntactic construct – namely, adjective–noun constructs – is hard to explain. Furthermore, a syntactic theory of Hebrew is bound to face difficulties in accounting for definite NPs. The same is true for a compositional semantic theory for Hebrew. We provided an HPSG analysis of Hebrew NPs, demonstrating that it is much easier to construct once definiteness is considered a morphological process, taking place in the lexicon. This work is part of a broader project whose aim is to provide an HPSG-based grammar for NPs in Hebrew. The analyses described herein are incorporated into this larger project.

Acknowledgments

This paper would never have been written without the help, encouragement and good advice of Paola Monachesi. I have benefited greatly from long discussions on these issues with Edit Doron and with Gerald Penn. I am grateful to Nissim Francez, Tali Siloni, Detmar Meurers, Hagit Borer, Bob Borsley and an anonymous CLIN reviewer for reading various drafts of this paper and commenting on them. I wish to thank Alon Itai for giving me access to on-line Hebrew corpora. The ideas expressed here, especially the erroneous ones, are all mine. This work was supported by the Minerva Stipendien Komitee.

[10] The elements that heads subcategorize for are SYNSEMs, not signs, so this information has to be explicitly encoded, but this is a technical minor problem.

References

Abney, Steven. 1987. *The English Noun Phrase in Its Sentential Aspect.* Ph.D. thesis, Massachusettes Institute of Technology, Cambridge.

Anderson, Stephen R. 1992. *A-morphous Morphology*, volume 62 of *Cambridge Studies in Linguistics.* Cambridge University Press, Cambridge.

Borer, Hagit. 1984. *Parametric Syntax - Case Studies in Semitic and Romance Languages*, volume 13 of *Studies in Generative Grammar.* Foris Publications, Dordrecht – Holland.

Borer, Hagit. 1988. On the morphological parallelism between compounds and constructs. In Geert Booij and Jaap van Marle, editors, *Yearbook of Morphology 1.* Foris publications, Dordrecht – Holland, pages 45–65.

Borer, Hagit. 1994. The construct in review. In Jacqueline Lecarme and Ur Shlonsky, editors, *Proceedings of the Second Conference on Afroasiatic Linguistics*, Sophia Antipolis, France, June. (to appear in *Studies in Afroasiatic Grammar*).

Börjars, Kersti E. 1992. An argument in favor of a syntax-morphology distinction: Scandinavian and Balkan noun phrases. In Laura A. Buszard-Welcher, Lionel Wee, and William Weigel, editors, *Proceedings of the eighteenth annual meeting of the Berkeley Linguistics Society*, pages 276–286, Berkeley, CA, February. BLS.

Borsley, Robert D. 1989. An HPSG approach to Welsh. *Journal of Linguistics*, 25:333–354.

Borsley, Robert D. 1995. On some similarities and differences between Welsh and Syrian Arabic. *Linguistics*, 33:99–122.

Di Sciullo, Anna Maria and Edwin Williams. 1987. *On the definition of word*, volume 14 of *Linguistic Inquiry Monographs.* MIT Press, Cambridge, MA.

Givón, T. 1981. On the development of the numeral 'one' as an indefinite marker. In Hagit Borer and Youssef Aoun, editors, *Theoretical Issues in the grammar of Semitic Languages*, volume 3 of *MIT Working Papers in Linguistics.* MITWPL, Cambridge, MA, pages 233–255.

Halpern, Aaron. 1992. The Balkan definite article and pseudo-second position. In Laura A. Buszard-Welcher, Lionel Wee, and William Weigel, editors, *Proceedings of the eighteenth annual meeting of the Berkeley Linguistics Society*, pages 338–349, Berkeley, CA, February. BLS.

Hazout, Ilan. 1991. *Verbal Nouns: Theta Theoretic Studies in Hebrew and Arabic.* Ph.D. thesis, University of Massachusetts, Amherst, MA 01003, February.

Hinrichs, Erhard W. and Tsuneko Nakazawa. 1996. Applying lexical rules under subsumption. In *Proceedings of COLING*, pages 543–549, Copenhagen.

Kolliakou, Dimitra. 1996. Definiteness and the make-up of nominal categories. In Claire Grover and Enric Vallduví, editors, *Studies in HPSG*, volume 12 of *Edinburgh Working Papers in Cognitive Science.* Centre for Cognitive Science, The University of Edinburgh, chapter 4, pages 121–164.

Levi, Judith N. 1976. A semantic analysis of Hebrew compound nominals. In Peter Cole, editor, *Studies in Modern Hebrew Syntax and Semantics*, number 32 in North-Holland Linguistic Series. North-Holland, Amsterdam, pages 9–55.

Miller, Philip. 1992. Postlexical cliticization vs. affixation: Coordination criteria. In Costas Canakis, Grace Chan, and Jeanette Denton, editors, *Papers from the 28th Regional Meeting of the Chicago Linguistic Society*, pages 382–396, Chicago.

Miller, Philip. 1993. Morphological marking misses the head. In Jonathan Mead, editor, *The Proceedings of the Eleventh West Coast Conference on Forlam Linguistics*, pages 341–353. Stanford Linguistics Association, Center for the Study of Language and Information.

Netter, Klaus. 1994. Towards a theory of functional heads. In John Nerbonne, Klaus Netter, and Carl Pollard, editors, *German in Head-Driven Phrase Structure Grammar*, volume 46 of *CSLI Lecture Notes*. CSLI, Stanford, CA, chapter 9, pages 297–340.

Ornan, Uzzi. 1964. *Noun Phrases in Modern Hebrew Literature*. Ph.D. thesis, Hebrew University, Jerusalem. (in Hebrew).

Ornan, Uzzi. 1994. Basic concepts in "romanization" of scripts. Technical Report LCL 94-5, Laboratory for Computational Linguistics, Technion, Haifa, Israel, March.

Pollard, Carl and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press and CSLI Publications.

Ritter, Elizabeth. 1988. A head-movement approach to construct-state noun phrases. *Linguistics*, 26(6):909–929.

Ritter, Elizabeth. 1991. Two functional categories in noun phrases: evidence from Modern Hebrew. In Susan D. Rothstein, editor, *Perspectives on Phrase Structure: Heads and Licensing*, volume 25 of *Syntax and Semantics*. Academic Press, pages 37–62.

Rosén, Haiim B. 1977. *Contemporary Hebrew*, volume 11 of *Trends in Linguistics – State of the Art Reports*. Mouton, The Hague.

Shlonsky, Ur. 1990. Hebrew construct state nominals, Arabic verb-initial clauses and the head movement constraint. Based on talk delivered at the parasession on Non-Indo European languages, GLOW colloquium, Cambridge, London.

Siloni, Tali. 1991. Noun raising and the structure of noun phrases. In Jonathan David Bobaljik and Tony Bures, editors, *Papers from the third Student Conference in Linguistics*, volume 14 of *MIT Working Papers in Linguistics*. Department of Linguistics and Philosophy, MIT, Cambridge, MA 02139, pages 255–270.

Siloni, Tali. 1994. *Noun Phrases and Nominalizations*. Ph.D. thesis, Départment de linguistique générale, Université De Genève.

Spencer, Andrew. 1991. *Morphological Theory*. Basil Blackwell, Oxford, UK.

Wintner, Shuly. 1998a. Definiteness agreement and inheritance in Hebrew. Paper presented at the 14th annual meeting of the Israeli Association for Theoretical Linguistics (IATL-14), Ben Gurion University, Beer Sheva, and at the 4th Conference on Afro-asiatic Languages, SOAS, London, June.

Wintner, Shuly. 1998b. Noun phrases as NPs – the case of Hebrew. In Gosse Bouma, Geert-Jan Kruijff, and Richard Oehrle, editors, *Proceedings of the 4th conference on Formal Grammar and the 5th conference on Head-driven Phrase Structure Grammar (FHCG'98)*, Saarbrücken, Germany, August.

Zwicky, Arnold M. 1977. On clitics. Reproduced by Indiana University Linguistics Club, Bloomington.

Zwicky, Arnold M. 1985. Clitics and particles. *Language*, 61(2):283–305.

Zwicky, Arnold M. and Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language*, 59(3):502–513.