Extending partial pro-drop in Modern Hebrew: A comprehensive analysis

Nurit Melnik
University of Haifa

Proceedings of the 14th International Conference on
Head-Driven Phrase Structure Grammar

Stanford Department of Linguistics and CSLI's LinGO Lab

Stefan Müller (Editor)

2007

Stanford, CA: CSLI Publications

pages 173–193

Melnik, Nurit. 2007. Extending partial pro-drop in Modern Hebrew: A comprehensive analysis. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford Department of Linguistics and CSLI's LinGO Lab*, 173–193. Stanford, CA: CSLI Publications.
DOI: 10.21248/hpsg.2007.11.

Abstract

Modern Hebrew is considered to be a 'partial *pro*-drop language'. Traditionally, the distinction between cases where *pro*-drop is licensed and those in which it is prohibited, was based on the person and tense features of the verb: 1st and 2nd person pronominal subjects may be omitted in past and future tense. This generalization, however, was found to be false in a number of papers, each discussing a subset of the data. Thus, contrary to conventional wisdom, dropped 3rd person pronouns subjects do occur in the language in particular contexts.

Identifying these contexts by way of a corpus-based survey is the initial step taken in this study. Subsequently, a careful syntactic analysis of the data reveals broad generalizations which have not been made to date. Thus, what was initially assumed to be a uniform phenomenon of 3rd person *pro*-drop turns out to be manifested in three distinct types of constructions. Finally, the proposed HPSG-based analysis incorporates insights concerning correlations between finite and non-finite control, non-canonical elements, locality, and binding.

1 Introduction

The phenomenon of *pro*-drop whereby pronominal arguments may be omitted in particular contexts is well-known and well-studied. Moreover, the notion of the Null Subject Parameter, which presumably distinguishes between those languages which allow unexpressed pronominal subjects (i.e., *pro*-drop languages) and those which do not, is prevalent in the transformational syntax literature. Modern Hebrew (MH) poses a challenge to this bifurcation since it exhibits what is referred to as 'partial *pro*-drop', where *pro*-drop is only partially licensed in the language.

Traditionally, the distinction between cases where *pro*-drop is licensed in MH, and those in which it is prohibited, was based on the person and tense features of the verb. This generalization, however, was shown to be empirically false in several papers (Borer 1989, Ariel 1990, Vainikka and Levy 1999, and Gutman 2004), each discussing a subset of the data, from one particular aspect.

In this paper I take a broader perspective by first conducting a comprehensive corpus-based survey[1] of cases in which the traditional distinction fails, followed by a careful syntactic analysis of the data. This process, as I show, reveals broad generalizations which have not been made to date, as well as insights concerning the correlation between the control of unexpressed subjects of infinitival complements and the identification of dropped subjects in finite complement clauses.

[†]This research was supported by the Israel Science Foundation (grant no. 137/06) and by The Caesarea Edmond Benjamin de Rothschild Foundation Institute for Interdisciplinary Applications of Computer Science. I am thankful to Shuly Wintner for his feedback and discussions and to the anonymous reviewers and the participants of HPSG 2007 for their comments.

[1]The Haaretz Corpus, compiled from a daily newspaper in Hebrew, was provided to me by the Knowledge Center for Processing Hebrew (http://mila.cs.technion.ac.il).

2 *Pro*-drop in Modern Hebrew

The licensing conditions of null pronominal subjects in MH is often attributed to the person and tense features. Thus, 1st and 2nd person pronominal subjects may be omitted in past and future tense (1). Overt pronouns in this context are used for emphasis or contrastively.

(1) (ata) **axalta**/**toxal** tapuax
(you) ate/will-eat.2SM apple
"You ate/will eat an apple."

Pro-drop is not possible with third person pronominals (2a) and in all cases of present tense, regardless of the agreement properties of the subject (2b).

(2) a. *(hu) **axal** tapuax
(He) ate.3SM apple
"He ate an apple."

b. *(ani) **oxel** tapuax
(I) eat.SM apple
"I eat an apple."

The distinction between the two cases is often ascribed to the "richness" of the morphology. Past and future tense verbs in 1st and 2nd person are morphologically marked for person, number, and gender, while present tense verbs and third person verbs in past and future tense are marked for number and gender, but not for person. Thus, it is the person agreement feature which enables the identification of the dropped subject.

However, despite traditional observations, 3rd person *pro*-drop (3P-PD) is not completely banned from the language.[2] Sentence (3), taken from the Haaretz corpus, illustrates a number of contexts in which 3P-PD can occur.

(3) be-mixtav be-anglit ileget she-**hefits** bekerev kol ha-ovdim
in-letter in-English broken that-distributed.3SM among all the-workers
ha-zarim hoda la-hem beit ha-malon al
the-foreigners thanked.3SM to-them house.M.CS the-hotel for
avodat-am ha-kasha ve-**hodi'a** she-**yirkosh**
work-POSS.3PM the-hard and-announced.3SM that-will-buy.3SM
la-hem kartisei tisa le-artsotei-hem mi-kasp-am
to-them tickets.CS flight to-countries-POSS.3PM from-money-POSS.3PM
"In a letter in broken English which it distributed among all the foreign workers, the hotel management thanked them for their hard work and announced that it will buy them plane tickets to their countries at their own expense."

[2]Note that I do not consider impersonal or non-referential uses of verbs in 3rd person as 3P-PD.

First, the verb *hefits* ('distributed') heads a non-subject relative clause in which the unexpressed pronominal subject refers to the matrix subject ('the hotel'). Second, the verb *yirkosh* ('will purchase') heads a subordinate clause which functions as the complement of the verb *hodi'a* ('announced'), which in itself appears to be subjectless.

3 Previous analyses of 3rd person *pro*-drop

The phenomenon of MH *pro*-drop has been discussed in numerous papers. However, as I came to realize, in many papers the existence of 3P-PD is not acknowledged (see, for example, Shlonsky (1997)). In what follows I briefly survey a number of analyses which do address 3P-PD.

Borer (1989), working in the transformational framework, distinguishes between 1st and 2nd *pro*-drop, where she posits that a phonologically empty *pro* occupies the subject position, and 3rd person *pro*-drop, which she claims is realized as an anaphoric AGR. 3P-PD is licensed when the embedded AGR is bound by an NP in a higher clause which assigns reference to the empty subject. Borer supports her claim by drawing parallels between "regular" anaphoric elements and 3P-PD. According to her, both anaphors and anaphoric AGRs cannot be bound by split antecedents. As evidence, she presents the following ungrammatical example, in which the agreement properties marked on the subjectless verb do not match those of either one of the matrix arguments.

(4) *Rina amra le-Ran she-**hiclixu** ba-bxina
Rina.F said.3SF to-Ran.M that-succeeded.3P in-the-test

"Rina told Ran that they succeeded in the test." (Borer (1989) ex. 55a)

Vainikka and Levy (1999) draw on the parallel behavior of Hebrew and Finnish with respect to *pro*-drop and propose a unified analysis for the two languages. They distinguish between the referential nature of 1st and 2nd person, on the one hand, and 3rd person on the other, and claim that the distinction has syntactic reflexes. *Pro*-drop is licensed whenever a referent is available. In 1st and 2nd person the referent is in the immediate conversational context; in embedded clauses with 3rd person *pro*-drop the referent is in the matrix clause. While the technical syntactic details proposed by Vainikka & Levy differ from those of Borer's, as far as I can tell, their empirical coverage is similar. Both analyses predict that 3P-PD is possible in complement clauses, as long as there is a matrix-argument antecedent.

Ariel (1990) takes a different perspective by considering 3P-PD in the context of her Accessibility Theory. Ariel proposes a type of an accessibility hierarchy for each of the factors involved in *pro*-drop. The anaphoric element, which is the verb, may have different degrees of "richenss" of agreement marking. Antecedents have different levels of salience, or prominence. Finally, there are varying degrees of cohesion between units in which anaphor and antecedent may appear.

To illustrate the difference between her approach and that of Borer (1989), she provides a counter-example to Borer's claim regarding the unavailability of split antecedents.

(5) Noga$_i$ bikra et Shimon$_j$ al ma'amaro ha-shovinisti
Noga.F criticized.3SF ACC Shimon.M on his-article the-chauvinistic
kshe-**nas'u**$_{i+j}$ li-yrushalayim
when-went.3P to-Jerusalem

"Noga criticized Shimon on his chauvinistic article when they went to Jerusalem." (Ariel (1990), chapter 6, ex. 5a)

Ariel attributes the difference in grammaticality to the type of verb used. Complements of *amar* ('said'), she claims, do not share the same degree of cohesion to the matrix verb than other sentential complements. Ariel, however, overlooks the fact that while sentence (5) does show a grammatical occurrence of split antecedents, its syntactic structure is not identical to (4), since the dropped subject in this case is the subject of an adverbial clause, not a complement clause. This, as I will subsequently show, makes a difference.

Gutman (2004) continues Ariel's line of inquiry by comparing the distribution of null subjects in Hebrew, Finnish, and Rumanian, a typical *pro*-drop language, and testing various salience and cohesion factors. She considers the effect of saliency in terms of grammatical functions, agents vs. non-agents, and animates vs. inanimates, and concludes that MH is less restrictive in the distribution of 3P-PD than Finnish, in that it allows non-subjects, non-agents, and inanimates to act as antecedents to dropped 3rd person subjects. In terms of cohesion, she claims that when the meaning is kept constant there is not observable contrast in MH between subordination and conjunction.

In conclusion, the different studies reviewed here suffer from a number of shortcomings. First, each of the studies addresses only some of the constructions and is based on a limited data set. Furthermore, I have shown cases where the authors do not make a clear distinction between the different constructions. This, as I will presently demonstrate, obscures the data and weakens the analysis. For these reasons the goals of the following sections are (i) to conduct a pre-theoretic corpus-based survey of 3P-PD, and (ii) to provide a comprehensive account of the data.

4 A closer look at the data

The starting point of the current analysis is identifying the syntactic constructions which license 3P-PD. A survey of examples cited in the literature as well as "naturalistic" corpus examples reveals four syntactic environments where 3P-PD is licensed: (i) adverbial clauses, (ii) non-subject relative clauses, (iii) complement clauses, and (iv) coordinated constructions. In what follows I will discuss each one in turn.

4.1 Adverbial clauses

Judging from the corpus data, *pro*-drop is the unmarked choice for 3rd person pronominal subjects of adverbial clauses in past or future tense. No 3P-PD was found in present tense. In the majority of the cases the antecedent is the matrix subject, yet antecedents with other grammatical functions were found as well. Consider, for example, sentence (6), where the antecedent is oblique, and sentence (5) above, where the antecedent is split between the subject and direct object.

(6) hu haya yoshev leyad-am$_i$ kol ha-layla
he was.3SM sit.present.SM next-to-them.3PM all the-night
kshe-**naflu**$_i$ le-mishkav...
when-fell.3PM to-bed

"He would sit next to them all night when they were ill..." (Ha'aretz Corpus)

The fact that adverbial clauses, which are adjoined to the main clause, constitute an appropriate context for 3P-PD is not surprising in light of Ariel's (1990) prediction regarding the level of cohesion that is required between the unit which hosts the dropped pronoun and that in which the antecedent occurs.[3]

4.2 Relative clauses

Non-subject relative clauses, too, are able to host 3P-PD. While this construction is not explicitly mentioned in the literature on MH *pro*-drop, a number of examples of it were found in the corpus. One such example is given in (3) and is repeated in abbreviated form in (7).

(7) be-mixtav she-**hefits**$_i$ bekerev ha-ovdim hoda la-hem
in-letter that-distributed.3SM among the-workers thanked.3SM to-them
beit ha-malon$_i$...
house.M.CS the-hotel...

"In a letter which it distributed among the workers, the hotel management thanked them..."

Relative clauses, too, function as adjuncts, and thus form cohesive units with the matrix clause. This cohesion is the enabling condition for the antecedent-dropped subject relationship.

[3] Note a parallel construction in English: *When asked to Join the party, Bill declined.*

4.3 Complement clauses

The case of complement clauses is not as straightforward as the previous ones. This was already hinted at in the discussion of Ariel's analysis, where she singles out a particular lexical item, *amar* ('said'), whose complement clauses form less cohesive units with their matrix clauses. It appears that not all complement clauses are created equal in terms of 3P-PD. In what follows I distinguish between three distinct cases.

Many MH verbs which take infinitival VP complements can also take finite clauses as complements. This class of verbs is further divided into two classes. The first class, to which I refer here as 'full control verbs', exhibits the same control pattern with both infinite and finite complements. Thus, when the subject of the finite clause is unexpressed, its referent is identified with the same matrix argument as in the infinitival case. An example is given in (8a), where the controller of the unexpressed subject is the indirect object *ha-ma'askikm* ('the employers').

The subject of the embedded clause, however, is not restricted to 3P-PD. Rather, it can be a pronominal, coindexed or not with the controller, or any lexical NP (8b). Furthermore, similarly to English control phenomena, this relationship carries over to denominal verbs as well (8c). Examples of subject control verbs in this category are *hivtiax* ('promise'), *kiva* ('hope'), and *hitsi'a* ('offer').

(8) a. ha-va'ad darash me-ha-ma'asikim$_i$
the-union demanded from-the-employers.PM
lashalem/she-**yeshalmu**$_i$ maskorot
to-pay.INF/that-will-pay.3PM salaries
"The union demanded from the employers to pay salaries."

b. ha-va'ad darash me-ha-ma'asikim$_i$
the-union demanded from-the-employers.PM
she-hem$_{i/j}$/ha-menahalim **yeshalmu** maskorot
that-they/the-managers will-pay.3PM salaries
"The union demanded from the employers that they/the managers pay salaries."

c. drishat ha-va'ad me-ha-ma'asikim$_i$
demand.CS the-union from-the-employers.PM
lashalem/she-**yeshalmu**$_i$ maskorot
to-pay.INF/that-will-pay.3PM salaries
"The union's demand from the employers to pay salaries"

Note that since finite verbs in Hebrew are morphologically marked in agreement with their subjects, the form of the verb indicates explicitly which is its antecedent (and can be manipulated to check alternatives). It should be added that present tense in this case is ungrammatical.

The second class of verbs is referred to here as 'semi-control verbs'. For this class, control is limited only to the infinitival case. Thus, while the controller of

the unexpressed subject of the infinitival VP is the matrix subject (9a), the subject of the embedded finite clause **cannot** be coindexed with the matrix subject, whether it is expressed or unexpressed (9b). Lexical NPs or unbound pronominals are acceptable (9c).

(9) a. ha-maxlaka ratsta **livnot** et ha-batim...
the-department.SF wanted.3SF to-build ACC the-houses
"The department wanted to build the houses..." (attested example)

b. *ha-maxlaka$_i$ ratsta she-hi$_i$/she-∅ **tivne**$_i$ et
the-department.SF wanted.3SF that-she/that-∅ will-build.3SF ACC
ha-batim...
the-houses

c. ha-maxlaka$_i$ ratsta she-ha-irya$_j$/she-hi$_j$
the-department.SF wanted.3SF that-the-municipality.SF/that-she
tivne$_j$ et ha-batim...
will-build.3SF ACC the-houses
"The department wanted the municipality to build the houses..."

Other members of this class are *tixnen* 'plan', *hiskim* 'agree', and *serev* 'refuse'.

The third class of verbs, referred to as 'finite control verbs', are verbs which only take finite clauses as complements. A 3P-PD embedded subject is obligatorily controlled by the matrix subject (10a). Split antecedents are impossible (cf. (4)). Moreover, present tense is ungrammatical. When not a 3P-PD, the embedded subject can be a pronominal or any lexical NP, on a par with full control verbs (10b).

(10) a. ha-xevra$_i$ hodi'a ki **hixlita**$_i$ al hafsakat
the-company.SF announced.3SF that decided.3SF on stopping
yitsur ha-memisim...
production the-solvents
"The company announced that it has decided to stop producing the solvents." (Ha'aretz Corpus)

b. ha-xevra$_i$ hodi'a ki hi$_{i/j}$/ha-va'ada **hixlita**
the-company.SF announced.3SF that she/the-committee decided.3SF
al hafsakat yitsur ha-memisim...
on stopping production the-solvents
"The company announced that it/the committee has decided to stop producing the solvents."

This class includes verbs of statement, such as *hitshir* ('claim'), *siper* ('tell'), and *hodi'a* ('announced'), which are widespread in the newspaper corpus I checked. Furthermore, it appears from the corpus that 3P-PD is the preferred option with this

type of verbs in this register. Closely associated with the newspaper register is the use of the complementizer *ki* ('that'), which is seldom used as an embedding complementizer in spoken language.[4]

To summarize, the licensing of 3P-PD in complement clauses depends on the verb type. The following table lists the different types of verbs discussed, along with information regarding their complementation patterns and the availability of 3P-PD.

Verb Type	VP*inf*	S*fin*	3P-PD
Finite Control Verbs	*	√	√
Full Control Verbs	√	√	√
Semi-Control Verbs	√	√	*
Infinitival VP only	√	*	*

Note that the "Infinitival VP only" category is included in the table for completeness. Verbs in this category, for example *nisa* ('try'), are not compatible with a finite complement clause, and are therefor not candidates for 3P-PD.

We can then conclude that 3P-PD is licensed in the finite complement clauses of two types of verbs: verbs which only take finite clauses as complements (i.e., finite control verbs) and a subset of verbs which take both infinitival VPs and finite clauses as complements (i.e., full control verbs).

4.4 Coordinated constructions

Many corpus examples of 3P-PD, as well as constructed examples in the literature, are instances of coordination, where a subjectless verb appears in the second conjunct. Alongside straightforward VP-CONJ-VP strings, there are many cases in which the second conjunct is preceded by an adverbial. Sentences such as (11) are considered by Ariel (1990) and Gutman (2004) as "conjoined sentences" with 3P-PD in the second conjunct.

(11) hayom noga$_i$ hitxila im shimon u-le-da'ati maxar
today Noga.F started.3SF with Shimon.M and-to-my-mind tomorrow
tatxil$_i$ im david
will-start.3SF with David.M

"Today Noga made a pass at Shimon and in my opinion tomorrow she will make a pass at David." (Ariel (1990), chapter 6, ex. 6a)

Note that this construction is not amenable to a simple VP-coordination analysis. The clause-initial adverbial *hayom* ('today') has scope only over the first conjunct, as it is contrasted with the adverbial *maxar* ('tomorrow') in the second conjunct.

[4]The complementizer *ki* is frequently used in a different sense, meaning 'because'.

A purely syntactic VP-coordination analysis, then, would have to assume a discontinuous VP constituent.

An additional example is the matrix clause of sentence (3), repeated here in abbreviated and slightly modified form as (12). The first conjunct in sentence (12) is an instance of "triggered inversion", where a non-subject dependent (a PP, in this case) appears clause-initially and triggers subject-verb inversion. The result is a VSO word order, where the subject comes between the verb and its complement, thus splitting the VP constituent.

(12) ba-mixtav [hoda la-hem beit ha-malon$_i$ al
in-the-letter thanked.3SM to-them house.M.CS the-hotel on
avodat-am] [ve-**hodi'a**$_i$ she-yirkosh la-hem
work-POSS.3PM and-announced.3SM that-will-buy.3SM to-them
kartisei tisa]
tickets.CS flight

"In the letter the hotel management thanked them for their work and announced that it will buy them plane tickets..."

One important characteristic which sets this construction from the previous ones is that the coordinate construction allows 3P-PD with a present tense verb in the second conjunct. This is illustrated in (13).

(13) asrot anashim$_i$ magi'im mi-tailand le-israel kshe-hem$_i$
tens.CS people arrive.PM from-Thailand to-Israel while-they
nirshamim ke-mitnadvim ax le-ma'ase **meshamshim**$_i$ ovdim
register.PM as-volunteers but actually serve.PM workers.PM
sxirim zolim
paid.PM cheap.PM

"Tens of people arrive from Thailand to Israel registered as volunteers while they actually work as low paid workers." (Ha'aretz Corpus)

The construction illustrated by (12) is similar to the Subject Gap in Finite clauses (SGF) coordination construction which is found virtually in all Germanic languages and marginally in English (Wunderlich 1988, Kathol and Levine 1993, Kathol 1999).[5]

(14) In den Wald ging der Jager und fin einen Hasen
into the forest went the hunter and caught a rabbit

"The hunter went into the forest and caught a rabbit."

The similarity between the MH construction and the SGF coordination construction, which is found in non-*pro*-drop languages, as well as the construction's

[5]I thank an anonymous reviewer for this suggestion.

compatibility with present tense suggest that the unexpressed subject in the second conjunct is an instance of some type of construction-specific gapping, and not *pro*-drop.

A different case of interaction between 3P-PD and coordination is discussed by Ariel (1990). This is illustrated by the example sentence in (15).

(15) noga dibra im shimon$_i$ yafe, ve-*(laxen) **ya'azor**$_i$ la li-sxov
Noga.F spoke to Shimon.M nicely and-so will-help.3SM her to-carry
et ha-mizvada
ACC the-suitcase

"Noga spoke nicely to Shimon, and (so) he will help her carry the suitcase."
(Ariel (1990), exx. 6c & 6eii)

Unlike the previously mentioned coordinated construction, the dropped subject of the verb in the second conjunct is not identified with the subject of the first conjunct. Rather, it is the indirect object which antecedes the missing subject. Consequently, a VP-coordination analysis is irrelevant. Moreover, as Ariel notes, the adverbial preceding the second conjunct is obligatory.

The role of the adverbial in licensing the 3P-PD in this case is creating cohesion between the two coordinated units by explicitly marking that the second clause is a consequence of the first. This is the type of construction referred to by Foley and Van Valin (1984) as 'cosubordination'.

To summarize, I propose that of all the coordinated constructions only those in which the dropped subject in the second conjunct is identified with an argument other than the subject are true cases of 3P-PD. Moreover, those are the cases where the obligatory occurrence of an adverbial subordinates the second conjunct to the main clause. In contrast, coordinated constructions where the subject of the first conjunct antecedes the empty subject in the second conjunct are instances of gapping.

4.5 Summary

At this point it has been established that contrary to conventional wisdom, 3rd person pronouns may be omitted in Modern Hebrew. Moreover, it has been shown that 3P-PD is licensed in a number of distinct constructions. One question remains, however, which is whether what we referred to here as 3P-PD is in fact "real" *pro*-drop.

In all the constructions in which they are licensed, dropped 3rd person pronominal subjects require linguistic antecedents. This characteristics sets them apart from "standard" *pro*-drop, which does not impose such a constraint. In Ariel's (1990) terms, the impoverished accessibility of 3rd person referents as identifiers of unexpressed subjects (in comparison with highly accessible 1st and 2nd person referents) requires there to be a linguistic antecedent in the matrix clause to identify the dropped 3rd person pronominal subject.

The tense restriction, which prohibits 1st and 2nd person *pro*-drop from occurring in present tense, applies to 3P-PD in adjunct clauses and complement clauses. Nevertheless, dropped 3rd person subjects in present tense coordinated constructions are grammatical. This, I claims, rules out the possibility of associating 3P-PD in coordinate constructions with *pro*-drop. This type of construction is similar to the SGF coordination constructions, which is also found in non-*pro*-drop languages (e.g,. German and English).

As to 3P-PD in adjunct and complement clauses, the main distinction between this type of subject drop and that of 1st and 2nd person is the nature of the licensing conditions. 1st and 2nd person *pro*-drop is licensed regardless of the syntactic construction in which it appears. In contrast, the distribution of 3P-PD is constrained by the type of syntactic construction. 3P-PD in adjunct clauses can be anteceded by a single or a split matrix antecedent. 3P-PD in complement clauses is licensed lexically by the embedding verb, and not by the verb whose pronominal subject is dropped. Moreover, the identification of the referent (or controller) of the unexpressed subject is lexically specified at the matrix verb level.

Consequently, I conclude that while there indeed are similarities between "standard" *pro*-drop and 3P-PD, the two phenomena cannot be conflated. Moreover, what was at first assumed to be a uniform phenomenon of 3P-PD has turned out to be manifested in three distinct types of constructions.

5 The proposed analysis

5.1 Overview

The main challenges which 3P-PD in Modern Hebrew poses are threefold: accommodating non-local constraints, accounting for the two types of dropped subjects, and providing an analysis of the different control patterns in complement clauses. In what follows I will undertake each of the challenges in the process of presenting an account of the phenomenon.

5.1.1 Non-local constraints

The 3P-PD constructions presented here raise issues regarding the locality of selection, in that they require that information regarding the subject of a finite clause be visible at the CP level. Thus, in all relevant constructions the licensing of 3P-PD does not occur at the lexical level, where the verb combines with its dependents, but rather, at the clausal level. This, of course, is problematic in a framework such as HPSG where valence requirements are canceled off as they are realized in the construction of phrasal signs. Once the SUBJ requirement is fulfilled it is assumed to be non longer on the VALENCE lists.

In this issue, Sag (2007) mentions similar cases of controlled pronominal subjects in finite clauses in the context of his discussion of locality. The solution which he proposes for such cases, as well as other related phenomena, is the category

feature EXTERNAL ARGUMENT (XARG). Unlike VALENCE requirements, which are cancelled off from the list as they are realized, the XARG feature percolates information "beyond" the phrasal level. As such, this feature provides a handle to information inside the clause, and thus overcomes the locality issue.

The visibility of the XARG feature at the clausal level enables us to define clausal constraints which target properties of the clausal subject. More specifically, this requires that the XARG feature percolate from the lexical level to the CP level. This, I propose, is achieved by the coindexation of the complementizers' XARG feature with the XARG of the clause which they select.

(16)
$$\left[\begin{array}{l}\text{SS}\,|\,\text{LOC}\,|\,\text{CAT}\left[\begin{array}{l}\text{HEAD } c\\ \text{XARG } \boxed{1}\\ \text{VAL}\,|\,\text{COMPS}\ \left\langle\left[\begin{array}{l}\textit{fin-clause}\\ \text{HEAD}\left[\begin{array}{l}\textit{verb}\\ \text{VFORM } \textit{fin}\end{array}\right]\\ \text{XARG } \boxed{1}\end{array}\right]\right\rangle\end{array}\right]\end{array}\right]$$

Overcoming the locality barrier is the first step in providing an analysis of 3P-PD in its various manifestations. The second step is to determine the exact nature of the unexpressed 3rd person pronominal subject, and to distinguish it from "standard" 1st and 2nd person *pro*-drop.

5.1.2 *Pro*-drop

The analysis of *pro*-drop in HPSG builds on the disassociation between ARG-ST and VALENCE proposed by Manning and Sag (1998). Thus, *pro*-drop is viewed as a variation on the Argument Realization Principle (ARP), where the least oblique argument in ARG-ST is not mapped to a VALENCE slot, yet remains in ARG-ST (Ginzburg and Sag, 2000). A preliminary version of the MH *pro*-drop ARP, which incorporates the language-specific tense & person restrictions and reflects the traditional description of *pro*-drop, is given in (17).

(17) *Pro*-drop ARP (preliminary version)
$$\left[\begin{array}{l}\text{HEAD}\left[\begin{array}{l}\textit{v}\\ \text{VFORM } \textit{past}\ \vee\ \textit{future}\end{array}\right]\\ \text{VAL}\left[\begin{array}{l}\text{SUBJ}\,\langle\rangle\\ \text{COMPS } \boxed{1}\end{array}\right]\\ \text{ARG-ST}\left\langle\text{NP:}\left[\begin{array}{l}\textit{ppro}\\ \text{PER } \textit{1st}\ \vee\ \textit{2nd}\end{array}\right]\right\rangle \oplus \boxed{1}\ \textit{list}\end{array}\right]$$

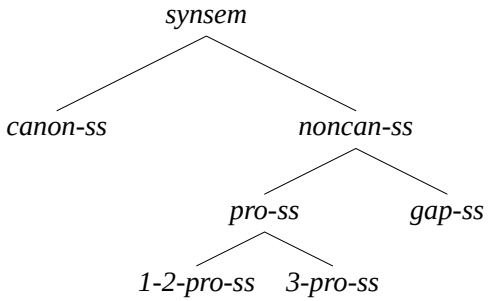
This type of constraint could suffice for the purpose of accounting for "standard" *pro*-drop in the language, since overt 1st and 2nd person pronoun subjects

can be freely omitted (modulo pragmatic considerations), regardless of the syntactic context, and consequently the status of a *pro*-dropped clause is identical to that of its overt-pronoun counterpart. 3P-PD, however, as was previously shown, has a much more restricted distribution. Moreover, the licensing conditions of 3P-PD target "higher" clauses, where the subject requirements of the lower verb are not visible. In other words, the fact that a 3rd person pronominal was dropped needs to be projected at the clausal level. For this reason the XARG feature should be incorporated into the *pro*-drop constraint. Moreover, the value of XARG should reflect the fact that the subject is "dropped" or unexpressed.

The HPSG type inventory provides a way to account for arguments which are not realized locally by overt linguistic expressions. These arguments are licensed by non-canonical synsems (*noncan-ss*), in distinction from canonical sysems (*canon-ss*), which license overt expressions. The type hierarchy given in (18), is an extension of the hierarchy posited by Ginzburg and Sag (2000). Ginzburg and Sag's hierarchy defines two subtypes of *noncan-ss*: *gap-ss*, which refers to 'gap' arguments in extraction constructions, and *pro-ss*, which accounts for unexpressed controlled subjects of nonfinite phrases.

For the purpose of this account I propose a slight extension. Under this analysis the use of *pro-ss* is extended to the domain of finite phrases, and, in addition, is further expanded by the introduction of two immediate subtypes: *1-2-pro-ss* and *3-pro-ss*. As will be shown, this architecture provides a way of both distinguishing and consolidating the two types of dropped subjects.

(18)



Consequently, the proposed *Pro*-drop Argument Realization Principle is given in (19). Note that the relationship between the unexpressed pronominal subject in ARG-ST and the non-canonical pronominal in XARG is maintained by the coindexation of the CONTENT value of the two features. Thus, once constructed, the phrase projects the INDEX feature of its unexpressed subject, as well as the information that it contains a non-canonical subject.

(19) *Pro*-drop ARP (final version)

$$\begin{bmatrix} \text{HEAD} \begin{bmatrix} v \\ \text{VFORM } \textit{past} \ \vee \ \textit{future} \end{bmatrix} \\ \text{VAL} \begin{bmatrix} \text{SUBJ} \langle\rangle \\ \text{COMPS } \boxed{1} \end{bmatrix} \\ \text{ARG-ST} \left\langle \text{NP: } \boxed{2} \textit{ppro} \right\rangle \oplus \boxed{1} \textit{ list} \\ \text{XARG } \textit{pro-ss}\text{: } \boxed{2} \end{bmatrix}$$

It should be added, for completeness, that in the "standard" ARP the XARG value is identified with that of (the first and only element of) SUBJ.

5.1.3 Adjunct clauses

The licensing of 3P-PD in adjunct clauses is defined in contrast to its prohibition in root clauses. Both constraints apply to clausal types. Following Sag (1997) and Ginzburg and Sag (2000), relative clauses are licensed by subtypes of the clausal type *rel-cl*. The distinguishing characteristics of all relative clauses are: (i) they cannot serve as independent clauses, (ii) they cannot be inverted, and (iii) the modify nominals. These characteristics are expressed by way of type constraints on the supertype *rel-cl*.

The aforementioned studies do not consider adverbial clauses. However, I assume that in addition to the *rel-cl* type an analogous type, *adv-cl*, is needed in order to account for adverbial clauses, which, similarly to relative clauses, (i) cannot serve as independent clauses, (ii) cannot be inverted, and (iii) have a non-empty MOD feature. Naturally, the MOD value of adverbials is not *noun*, but *v*. The question of whether *rel-cl* and *adv-cl* are subtypes of a more general type (*mod-cl*) is immaterial to the present analysis. The crucial issue is that both types of clauses allow their XARG value to be of type *3-pro-ss*. In contrast, clauses which function as root clauses are incompatible with a *3-pro-ss* XARG. This generalization can be captured either by a default constraint on all clauses, or explicitly on the most general clause types which function as root clauses. An illustration of an analysis of 3P-PD in an adverbial clause (extracted from (5)) is given in figure 1.

Recall that the identification of the referent of the unexpressed embedded subject depends on linguistic antecedents in the matrix clause. This, however, is a pragmatic process, which is not syntactically determined, and, thus permits both single or split antecedents.

5.1.4 Complement clauses

As was previously discussed, the control patterns involved with 3P-PD are quite complex. An account of these patterns is required to distinguish between three different verb categories: *full control verbs*, *semi-control verbs*, and *finite control verbs*. In what follows I address each one in turn.

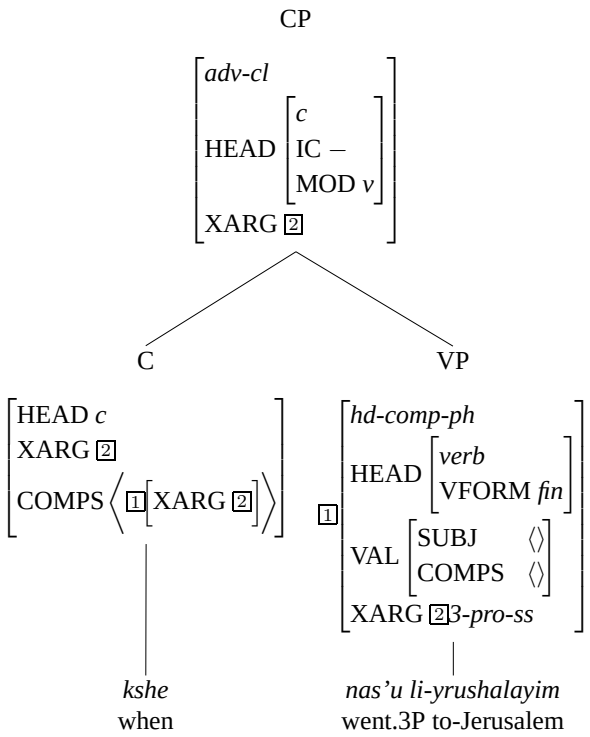


Figure 1: Adverbial Clause

Full control verbs

The class of *full control verbs* is the least restrictive one. Verbs which belong to this class alternate between taking infinitival and finite clauses as complements. The infinitival case is remarkably similar to that of English, and, therefore compatible with the analysis proposed by Pollard and Sag (1994). Verbs fall into two categories — subject control and object control — according to the grammatical function of the matrix argument which controls the unexpressed subject of the VP complement. Control in this case is obligatory.

Finite control is more involved. The subject of the finite complement clause may not necessarily be controlled by a matrix argument. Thus, as was illustrated in (8) above, the embedded subject can be a controlled *3-pro-ss*, a controlled or free personal pronoun (*ppro*), or an unbound lexical NP (*npro*). An additional complication, not mentioned earlier, is the possibility of the occurrence of an uncontrolled *1-2-pro-ss*. An example is given in (20).

(20) ha-va'ad darash me-ha-ma'asikim$_i$ she-**neshalem**$_i$ maskorot
the-union demanded from-the-employers.PM that-will-pay.1P salaries

"The union demanded from the employers that we pay salaries."

In order to capture the different patterns, I propose to differentiate between

those cases in which control is obligatory and those in which it is not. Consequently, a lexical rule will account for the control pattern correspondence between the infinitival and the finite cases. The Infinite to Finite Subject Control Lexical Rule for subject control verbs such as *hivtiax* ('promise') is given in (21).

(21) Infinite to Finite Subject Control Lexical Rule

$$\begin{bmatrix} \textit{inf-subj-full-ctrl} \\ \text{CAT} \mid \text{VAL} \begin{bmatrix} \text{SUBJ} \langle \text{NP}\boxed{1} \rangle \\ \text{COMPS} \left\langle \text{VP} \begin{bmatrix} \text{VFORM } \textit{inf} \\ \text{SUBJ} \langle \text{NP}\boxed{1} \rangle \end{bmatrix} : \boxed{3} \right\rangle \end{bmatrix} \\ \text{CONTENT} \begin{bmatrix} \text{RELATION } \textit{rel} \\ \text{ARG1 } \boxed{1} \\ \text{SOA-ARG } \boxed{3} \end{bmatrix} \end{bmatrix} \Rightarrow \begin{bmatrix} \textit{fin-subj-ctrl} \\ \text{CAT} \mid \text{VAL} \begin{bmatrix} \text{SUBJ} \langle \text{NP}\boxed{1} \rangle \\ \text{COMPS} \left\langle \begin{bmatrix} \textit{fin-clause} \\ \text{HEAD } c \\ \text{XARG } \textit{pro-ss}\boxed{1} \end{bmatrix} : \boxed{3} \right\rangle \end{bmatrix} \\ \text{CONTENT} \begin{bmatrix} \text{RELATION } \textit{rel} \\ \text{ARG1 } \boxed{1} \\ \text{SOA-ARG } \boxed{3} \end{bmatrix} \end{bmatrix}$$

It should be emphasized that the ability to "look inside" the finite complement is achieved by way of the XARG feature which exposes the type of subject and its CONTENT value. The structure-sharing of index features, indicated by $\boxed{1}$, renders the control obligatory. A similar rule is required for object control verbs.

A partial analysis of the finite object control example in (8a) is given in figure 2.

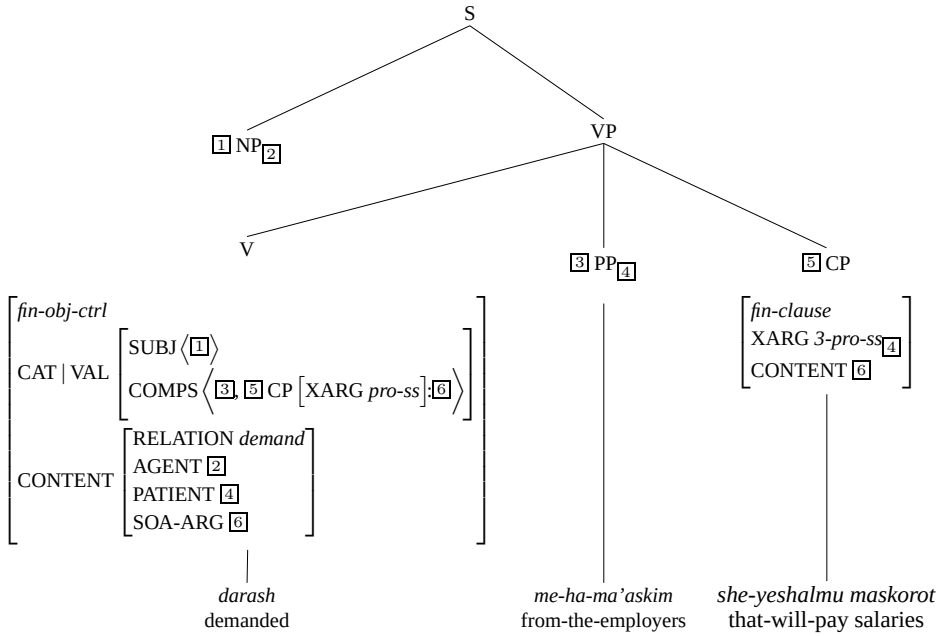


Figure 2: Finite Object Control

The remaining cases are those in which the embedded subject is not necessarily

coindexable with the matrix subject. A description of the associated lexical type is given in (22).

(22)
$$\left[\begin{array}{l} \textit{finite-comp} \\ \text{CAT} \mid \text{VAL} \left[\begin{array}{ll} \text{SUBJ} & \left\langle \text{NP}\boxed{1} \right\rangle \\ \text{COMPS} & \left\langle \left[\begin{array}{l} \textit{fin-clause} \\ \text{HEAD } \textit{c} \\ \text{XARG} \left[\begin{array}{l} \textit{canon-ss} \\ \text{HEAD } \textit{noun} \end{array}\right] \vee \textit{1-2-pro-ss} \end{array}\right] : \boxed{3} \right\rangle \end{array}\right] \\ \text{CONTENT} \left[\begin{array}{l} \text{RELATION } \textit{rel} \\ \text{ARG1 } \boxed{1} \\ \text{SOA-ARG } \boxed{3} \end{array}\right] \end{array}\right]$$

As was previously mentioned, different types of nominal subjects can serve as embedded subjects in this construction. This is expressed in the XARG value of the finite clause in the COMPS list. Nominal external arguments of type *cannon-ss* account for lexical NPs as well as pronominal ones. It should be noted that while the constraints do not impose a coindexation relation between the XARG and the SUBJ, they do not prevent it. Consequently, embedded pronominal subjects are either bound or free. The second disjunct in the XARG value is necessary in order to allow cases of 1st or 2nd person *pro*-drop in the complement clause, such as (20) above.

The use of disjunction in this constraint is not trivial with respect to the formalism of HPSG. However, the proposed type hierarchy of *synsems* does not allow for a natural grouping of these NPs (i.e., lexical NPs, personal pronouns, and 1st and 2nd person *pro*-drop). For the purpose of descriptive adequacy I choose to use the disjunction operator. An alternative solution is to posit different lexical entries for each of the XARG possibilities.

Semi-control verbs

The Infinite to Finite Subject Control Lexical Rule given in (21) does not apply to the class of semi-control verbs, since control in this case is restricted to the nonfinite domain, similarly to English control verbs. Thus, while as infinitival control verbs the two types of verbs are indistinguishable, the types which license them must be distinct. For this reason I posit two separate types, *inf-subj-full-ctrl* and *inf-subj-semi-ctrl*, which are both subtypes of more general type *inf-subj-ctrl*.

In addition to infinitival VPs, finite clauses too can serve as complements to semi-control verbs, provided that the embedded subject is not controlled by the matrix subject (see (9b) & (9c)). This completely rules out any type of NP, canonical or non-canonical, which is coindexed with the subject.

One way to build this type of a constraint into the grammar is by using inequation, and stating that the indices of the two entities cannot be coindexed. This, of

course, raises the issue of the status of inequation in the formalism of HPSG, a debate which is not the focus of this paper. An alternative option is to associate this constraint with Binding Theory. More specifically, according to Principle B, a personal pronoun must be a-free, where 'a-free' refers to the locus of the HPSG Binding Theory, namely ARG-ST (Manning and Sag, 1998). Since both overt pronouns and *pro*-sss are pronouns, the binding of XARG by the matrix subject can be avoided by adding it to the ARG-ST of the embedding clause.[6]. In such a configuration, XARG is in the binding domain of the subject, and thus cannot be coindexed with it.

(23)
$$\left[\begin{array}{l} \textit{finite-comp-no-bind} \\ \text{CAT}\left[\begin{array}{l} \text{VAL}\left[\begin{array}{l} \text{SUBJ}\left\langle \boxed{1}\,\text{NP}_{\boxed{4}}\right\rangle \\ \text{COMPS}\left\langle \boxed{2}\left[\begin{array}{l}\textit{fin-clause}\\ \text{HEAD } c\\ \text{XARG } \boxed{3}\text{NP}\end{array}\right]:\boxed{5}\right\rangle \\ \text{ARG-ST}\left\langle \boxed{1},\boxed{2},\boxed{3}\right\rangle \end{array}\right] \\ \text{CONTENT}\left[\begin{array}{l}\text{RELATION } \textit{rel}\\ \text{ARG1 } \boxed{4}\\ \text{SOA-ARG } \boxed{5}\end{array}\right] \end{array}\right] \end{array}\right]$$

At this point I consider the two alternatives as engineering solutions. I leave the question of the theoretical and empirical ramifications of each option to further research.

Finite control verbs

Finally, finite control verbs can only take finite clauses as complements. In fact, the type of constructions in which these verbs are licensed is a subset of those which license full subject control verbs, namely the finite ones. Consequently, the two lexical types which describe the realization possibilities of these verbs are *fin-subj-ctrl* (21) and *finite-comp* (22).

5.1.5 Coordinated constructions

The discussion of the coordinate constructions involved with 3P-PD distinguished between two types of constructions: an SGF-like construction, in which the unexpressed subject of the second conjunct is considered to be a gap, and cosubordination, where the unexpressed subject of the cosubordinated clause is identified with a non-subject in the first clause. An analysis of these constructions is outside the scope of this paper and is left for future work. Nevertheless, an HPSG-based analysis of the SGF coordination construction is proposed by Kathol (1999) in

[6]Note that this move is possible due to the disassociation between ARG-ST and VALENCE proposed by Manning and Sag (1998)

a linearization framework, where linear order is considered conceptually distinct from constituent relations. In addition, a discourse functional analysis of SGF coordination in LFG is proposed by Frank (2002).

6 Conclusion

Contrary to the traditional description of *pro*-drop in MH, *pro*-drop of 3rd person pronouns does occur. Its distribution, however, is more restricted than that of 1st and 2nd person pronouns. The observation presented here is that 3P-PD occurs freely in adjunct subordinate clauses (i.e., adverbial clauses, relative clauses, and 'cosubordinated' clauses) when it is anteceded by a matrix argument antecedent (single or split). Cases which were previously viewed as 3P-PD in conjoined sentences were analyzed here as cases of gapping and not *pro*-drop. Consequently, it was proposed that the licensing of this kind of 3P-PD is associated with types of *clausal constructions*. Furthermore, the clausal association confirms Ariel's prediction regarding the necessity of cohesion between the units of the antecedent and dropped subject.

More restrictive licensing conditions were found to apply to embedded complement clauses, where the licensing of 3P-PD depends on lexical properties of the embedding verb. Three types of verbs were identified, each with its particular complementation and control patterns. For one type of verbs referred to as 'full control verbs' the identification of the antecedent of the empty subject was found to correlate with the identification of the controller of parallel constructions with an infinitival complement. More generally, the licensing of 3P-PD in complement clauses was found to be determined at the lexical level.

In conclusion, this study provided a comprehensive data-driven account of the phenomenon of 3P-PD, a phenomenon that has not received an adequate analysis up until now. The proposed HPSG-based analysis incorporated insights concerning locality, clausal vs. lexical constraints, correlations between finite and non-finite control, non-canonical elements, and binding.

References

Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.

Borer, Hagit. 1989. Anaphoric AGR. In Osvaldo Jaeggli and Kenneth J. Safir (eds.), *The Null Subject Parameter*, pages 69–109, Dordrecht: Kluwer.

Foley, William A. and Van Valin, Jr., Robert D. 1984. *Functional syntax and universal grammar*. Cambridge, England: Cambridge University Press.

Frank, Anette. 2002. A (Discourse) Functional Analysis of Asymmetric Coordination. In *Proceedings of the 7th International Lexical Functional Grammar Conference (LFG'02), July 3-5*, Athens, Greece, in October.

Ginzburg, Jonathan and Sag, Ivan A. 2000. *Interrogative Investigations: the form, meaning, and use of English interrogatives*. Stanford, California: CSLI Publications.

Gutman, Einat. 2004. Third person null subjects in Hebrew, Finnish and Rumanian: an accessibility-theoretic account. *Journal of Linguistics* 40, 463–490.

Kathol, Andreas. 1999. Linearization vs. phrase structure in German coordination constructions. *Cognitive Linguistics* 10(4), 303–342.

Kathol, Andreas and Levine, Robert D. 1993. Inversion as a linearization effect. In *Proceedings of NELS 23*, pages 207–221.

Manning, Chris and Sag, Ivan A. 1998. Argument structure, valence, and binding. *Nordic Journal of Linguistics* 21, 107–144.

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

Sag, Ivan A. 1997. English Relative Clause Constructions. *Journal of Linguistics* 33(2), 431–484.

Sag, Ivan A. 2007. Remarks on Locality. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar*, Stanford: CSLI Publications.

Shlonsky, Ur. 1997. *Clause Structure and Word Order in Hebrew and Arabic: An Essay in Comparative Semitic Syntax*. Oxford Studies in Comparative Syntax, New York: Oxford University Press.

Vainikka, Anne and Levy, Yonata. 1999. Empty Subjects in Hebrew and Finnish. *Natural Language and Linguistic Theory* 17, 613–671.

Wunderlich, Dieter. 1988. Some problems of coordination in German. In U. Reyle and C. Rohrer (eds.), *Natural language parsing and linguistic theories*, pages 289–316, Dordrecht: Reidel.