Relational-realizational parsing

Tsarfati, R.; Sima'an, K.

[Link to publication](#)

Relational-Realizational Parsing

Reut Tsarfaty and Khalil Sima'an
Institute for Logic, Language and Computation, University of Amsterdam
Plantage Muidergracht 24, 1018TV, Amsterdam, The Netherlands
{rtsarfat,simaan}@science.uva.nl

Abstract

State-of-the-art statistical parsing models applied to free word-order languages tend to underperform compared to, e.g., parsing English. Constituency-based models often fail to capture generalizations that cannot be stated in structural terms, and dependency-based models employ a 'single-head' assumption that often breaks in the face of multiple exponence. In this paper we suggest that the position of a constituent is a *form* manifestation of its grammatical *function*, one among various possible means of realization. We develop the *Relational-Realizational* approach to parsing in which we untangle the *projection* of grammatical functions and their means of *realization* to allow for phrase-structure variability and morphological-syntactic interaction. We empirically demonstrate the application of our approach to parsing Modern Hebrew, obtaining 7% error reduction from previously reported results.

1 Introduction

Many broad-coverage statistical parsers to date are constituency-based with a Probabilistic Context-Free Grammar (PCFG) or a Stochastic Tree Substitution Grammar (STSG) at their backbone. The majority of such models belong to a *Head-Driven* paradigm, in which a head constituent is generated first, providing a positional anchor for subsequent (e.g., Markovian) sisters' generation.

Constituency-based models, lexicalized and unlexicalized alike, demonstrate state-of-the-art performance for parsing English (Charniak, 1997; Collins, 2003; Klein and Manning, 2003; Bod, 2003), yet a direct application of such models to parsing less configurational languages often fails to yield comparable results. The parameters of such parsers capture generalizations that are easily stated in structural terms (e.g., subjects linearly precede predicates, VPs dominate objects, etc.) which may not be adequate for parsing languages with less configurational character.

A different vein of research explores data-driven dependency-based parsing methods (e.g., (McDonald et al., 2005)) which seem to be intuitively more adequate for the task. It turns out, however, that even such models fail to provide the desired remedy. Recent reports by (Nivre, 2007) delineated a class of richly-inflected languages with relatively free word-order (including Greek, Basque, and Modern Standard Arabic) for which the parsers performed poorly, regardless of the parsing method used. The need for parsing methods that can effectively cope with such phenomena doesn't seem to have been eliminated by dependency parsing — perhaps quite the contrary.

The essential argument we promote here is that in order to deal with the kind of variation that is empirically observed cross-linguistically an alternative view of the generation process is required. Our *Relational-Realizational* parsing proposal, strongly inspired by Relational Grammar (Perlmutter, 1982), takes grammatical relations such as 'Subject' and 'Predicate' as central, primitive notions of the syntactic representation, and retains a distinction between the projection of such relations and the means by which they are realized. The grammar we develop here, formally

represented as a PCFG, articulates two alternating generation phases: a *Relational* phase, in which a clause-level category projects a monostratal Relational Network (RN) representation for the clause, and a *Realizational* phase, in which the projected relations are realized in a certain surface configuration. Paradigmatic morphosyntactic representations are constructed for all non-terminal nodes, allowing for morphosyntactic interaction at various levels of the syntactic parse tree.

We illustrate the application of our theoretical reconstruction to the representation of clause-level categories in Modern Hebrew (MH) and their interaction with a handful of morphological features. The treebank grammar resulting from our application yields 13% error reduction relative to a treebank PCFG which uses the same information in the form of state-splits, and our best result shows a 7% error reduction over the best parsing results for MH so far. Through a quantitative and qualitative analysis we illustrate the advantages of the *Relational-Realizational* approach and its potential promise for parsing other "exotic" languages.

2 Background

Recent decades have seen a surge of interest in statistical models using a body of annotated text for learning the distributions of grammatically meaningful structures, in order to assign the most likely ones to unseen sentences. Probabilistic Context Free Grammars (PCFGs) have become popular in the articulation of such models, and unlexicalized treebank grammars (or representational variations thereof) were shown to perform reasonably well on English benchmark corpora (Johnson, 1998; Klein and Manning, 2003).

A major leap in the performance of PCFG-based statistical parsers has been introduced by the move towards a *Head-Driven* paradigm (Collins, 2003; Charniak, 1997), in which syntactic categories are enriched with head information percolated up the tree. The head-driven generation process allows one to model the relation between the information content of a constituent and the information content of its head-marked sister. At the same time, such models introduce a bias with respect to the positioning of a non-head constituent relative to its head-marked sister. The vast improvement in parsing results came about not without modeling costs, e.g., additional ad-hoc modifications for capturing complex structures such as conjunction.

An inherent difficulty with the application of constituency-based parsing models is the implicit assumption that the relation between the position of a constituent and its grammatical function is fully predictable. For languages with relatively free word-order, this assumption often breaks down. Distinguishing, e.g., "left" and "right" distributions for constituents of the same 'sort' implicitly takes the position of a constituent to be a primitive syntactic notion, and their grammatical function to be a secondary, derived one. Theoretical accounts show that this may be insufficient (Perlmutter, 1982). A subsequent difficulty with the head-driven paradigm, also shared by dependency-based parsing methods, is the stipulation that all grammatically relevant properties of a phrase are recovered from a single head. In fact, it is typologically established that grammatically meaningful properties of a constituent may jointly emerge from different surface forms dominated by it (*co-heads* or *multiple exponence* (Zwicky, 1993; Blevins, 2008)).[1]

The task we undertake here is to suggest a statistical generative parsing method which is linguistically plausible as well as technologically viable for parsing languages with relatively free word-order and variable means of realization. In what follows we remain within the computationally efficient framework of PCFGs, and propose a variation that draws on insights from syntactic and morphological theories that have been explored cross-linguistically.

3 Approach

3.1 Relational Grammars (RGs)

Relational Grammars (RGs) were introduced in the early 80's when attempts to find a universal definition for notions such as a 'Subject' in terms of various "behavioral properties" seemed to have failed (Perlmutter, 1982). The unsuccessful attempts to recover an adequate definition of grammatical functions in structural terms led to a revival of a view in which grammatical relations such as 'Subject' and 'Object' are primitive notions by

---

[1] We refrain here from referring to the increasingly popular approach of discriminative parsing, firstly, because we are interested in a generative parsing model that assigns a probability distribution to all sentence-structure pairs in the language, essentially allowing it to be used as a language model (e.g., in SR or SMT applications). Secondly, so far the features that have been explored in these frameworks are mainly those easily stated in structural terms, with not much effort towards modeling morphosyntactic interactions systematically.

which syntactic structures are defined (Postal and Perlmutter, 1977). This view proved useful for descriptive purposes, influencing the design of formalisms such as Arc-Pair grammars (Postal, 1982) and LFG (Bresnan, 1979).

The two main primitive elements used in RGs are (a) a set of nodes representing linguistic elements (which we refer to using upper case letters), and (b) a set of names of grammatical relations (which we refer to as $gr_1...gr_n$). RGs represent the fact that a linguistic element bears a certain relation to another element using a structure called an 'Arc', represented as $[gr_i(A, B)]$. Arcs are represented as arrows, with A the head of the Arc and B its tail, and a Relational Network (RN) is defined to be a set of Arcs that share a single head.[2]

Now, a few theoretical observations are due. Firstly, the essential difference between RGs and dependency-based grammars is that RNs take the linguistic element at the head of a network to be a clause-level category, not a particular surface form. The corresponding tails are then the various nominals bearing the different grammatical relations to the clause (including a 'Predicate', a 'Subject', an 'Object', etc.). In addition, RNs abstract away from elements such as auxiliary verbs and particles which do not have their own arc representation. RGs also differ from phrase-structure grammars in that their RNs are unordered. Therefore, linear precedence need not play a role in stating generalizations. RGs differ from both constituency- and dependency-based formalisms in that they do not weigh heavily the 'single-head' assumption – RNs may delineate a whole chunk as bearing a certain grammatical relation to the clause.

The set-theoretic notion of RNs in RGs abstracts away from surface phenomena crucial for generating phrase-structure trees. Thus, we next turn to modeling how grammatical relations are realized.

3.2 Form, Function and Separation

Morphological phenomena such as suprasegmentation, interdigitation, reduplication, subtractive morphology, templatic morphology, and methathesis demonstrate that it is sometimes impossible to find a direct correspondence between a certain part of a word (a 'morpheme') and the function it has in altering the word's meaning (Anderson, 1992). Attempts to model such morphological phenomena brought forward the hypothesis that 'form' and 'function' need not stand in one-to-one correspondence, and that one is not necessarily immediately predicted by the other. This hypothesis is known as the "Separation Hypothesis" (Beard, 1988). The problem of modeling certain surface phenomena then boils down to modeling form and function correlations, bearing in mind that these may be quite complex.

Bringing this general notion of separation into the syntactic derivation, we propose to view the position of a constituent in a phrase as its *form* and the articulated grammatical relation as its *function*. The task of learning the position of different constituents realizing the grammatical relations in an RN is now delegated to a statistical component. A set of parameters which we refer to as "configuration" determines the syntactic *position* in which each of the grammatical relations is to be realized.

3.3 Morphosyntactic Representations

In order to connect the abstract RN representation with the constituents that syntactic parse trees are "made of" we propose to view the internal nodes of a tree as *Morphosyntactic Paradigms*. Our morphosyntactic representation for constituents, loosely inspired by (Anderson, 1992), is a structured representation of morphological and syntactic properties for an internal node in the parse tree. In our model, the morphological features associated with a syntactic constituent are percolated from its dominated surface forms, and we allow the specification of head (PoS tag) information and structural features such as vertical markovization. Given the grammatical relation an element bears to a clause, it is statistically feasible to learn the morphosyntactic paradigm by which it is realized.

4 The Model

4.1 The Generative Model

Let $S_p \rightarrow \langle S_{c_1}\text{-}gr_1 \ldots S_{c_n}\text{-}gr_n \rangle$ be a context-free rule where S_p is the morphosyntactic representation of a parent constituent, $gr_1...gr_n$ are the grammatical relations forming its RN, and $\langle S_{c_1} \ldots S_{c_n} \rangle$ are ordered morphosyntactic representations of the child constituents bearing the respective relations to the parent. Our grammar then conceptualizes the generation of such a rule in three phases:

---

[2] RGs also define the notion of a stratum, a single level of syntactic representation, and for the current discussion we assume a monostratal representation. We do not claim that our framework is capable of dealing with the full range of phenomena multistratal RNs were shown to account for, yet there is nothing in our proposal that excludes extending the representation into a multistratal framework that does so.

- **Projection:**
  $S_p \rightarrow \{gr_i\}_{i=1}^n @S_p$
- **Configuration:**
  $\{gr_i\}_{i=1}^n @S_p \rightarrow \langle gr_1 @S_p \ldots gr_n @S_p \rangle$
- **Realization:**
  $\{gr_i @S_p \rightarrow S_{c_i}\}_{i=1}^n$

In the projection stage we generate the set of grammatical relations in the RN of a constituent. In the configuration stage we order these grammatical relations, and in realization we generate the morphosyntactic representation of each child constituent given the relation to its parent. Figure (1) shows the application of this process to two clauses bearing identical RNs that are in turn realized in different possible configurations.

This three-step process does not generate functional elements (such as auxiliary verbs and special-purpose particles) that are outside of constituents' RNs. We thus let the configuration stage place obligatory or optional "realizational slots" between the ordered elements (marked $gr_i : gr_j$), signalling periphrastic adpositions and/or modification. Note that modification may introduce more than one constituent, to be generated in realization.

- **Projection:**
  $S_p \rightarrow \{gr_i\}_{i=1}^n @S_p$
- **Configuration:**
  $\{gr_i\}_{i=1}^n @S_p \rightarrow$
  $\langle gr_0 : gr_1 @S_p \ gr_1 @S_p \ldots gr_n : gr_{n+1} @S_p \rangle$
- **Realization:**
  $\{gr_i @S_p \rightarrow S_{c_i}\}_{i=1}^n$
  $\{gr_i : gr_{i+1} @S_p \rightarrow \langle S_{c_{i_1}} \ldots S_{c_{i_{m_i}}} \rangle\}_{i=0}^n$

In figure (2), the configuration stage reserves a slot for an obligatory punctuation mark at the end of an affirmative sentence. It further reserves a slot for an optional adverbial modifier at a position commonly employed in MH for interjections.

In the current framework, grammatical relations may be realized in a certain surface position via configuration, or using explicit morphological marking per grammatical relation independently of linear context. Figure (3) demonstrates how the realization phase models the correlation between grammatical relations and morphological information percolated from dominated surface forms. In particular, our model can capture the interaction between marked features, e.g., the 'exclusive or' relation between definiteness and accusativity in marking direct objects in MH (Danon, 2001).

Finally, a conjunction structure in our model is simply an RN representation of multiple morphosyntactically equivalent conjuncts, as illustrated in figure (4). This modeling choice avoids the need to stipulate a single head for such structures (cf. head-driven processes) and allows the different conjuncts to share a realization distribution — essentially implying homogeneity in the assignment of heads and morphosyntactic features across conjuncts.

4.2 The Probabilistic Model

Our probablistic model is a PCFG, where CFG rules capture the three stages of generation. Every time we apply our projection-configuration-realization cycle we replace the rule probability with the probabilities of the three stages, multiplied ($n + \sum_{i=0}^n m_i$ daugthers , $gr_0 = gr_{n+1} =$null).

$$P(\langle S_{c_{i_1}}, .., S_{c_{i_{m_i}}} S_{c_i} \text{-} gr_i, S_{c_{i+1_1}}, .., S_{c_{i+1_{m_{i+1}}}} \rangle_{i=0}^n | S_p) =$$
$$P(\{gr_i\}_{i=1}^n | S_p) \times$$
$$P(\langle gr_0 : gr_1, g_1, \ldots \rangle | \{gr_i\}_{i=1}^n, S_p) \times$$
$$\prod_{i=1}^n P(S_{c_i} | gr_i, S_p) \times$$
$$P(\langle S_{c_{0_1}}, \ldots, S_{c_{0_{m_0}}} \rangle | gr_0 : gr_1, S_p) \times$$
$$\prod_{i=1}^n P(\langle S_{c_{i_1}}, \ldots, S_{c_{i_{m_i}}} \rangle | gr_i : gr_{i+1}, S_p)$$

The multiplication implements the independence assumption between form and function underlying the Separation Hypothesis, and the conditioning we articulate captures one possible way to model a systematic many-to-many correspondence.

4.3 The Grammar

We use a probabilistic treebank grammar in which the different parameters and their probability distributions are read off and estimated from the treebank trees. Clause-level (or clause-like) constituents such as S, SQ, FRAG, FRAGQ, internally complex VPs and a small number NPs can head RNs. For the rest we use flat CFG rules. We use a limited set of grammatical relations, namely, 'Predicate', 'Subject', 'Object' and 'Complement' — making the distinction between a nominal complement and a verbal (infinitival) one. Our linguistic elements are morphosyntactic representations of labeled non-terminal constituents, where the morphosyntactic representations of constituents incorporate morphological information percolated from surface forms and syntactic information about the constituent's environment.

(a)

S

NP-*SBJ* VP-*PRD* NP-*OBJ*

S

{*PRD,OBJ,SBJ*}@S

SBJ@S *PRD*@S *OBJ*@S

NP VP NP

(b)

S

VP-*PRD* NP-*SBJ* NP-*OBJ*

S

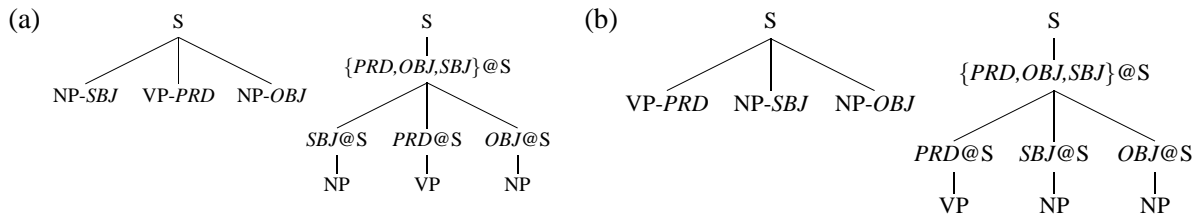{*PRD,OBJ,SBJ*}@S

PRD@S *SBJ*@S *OBJ*@S

VP NP NP

Figure 1: **Generating Canonical and Non-Canonical Configurations:** The CF depictions of the S level constituents at the LHS of (a) and (b) are distinct, whereas the *RR*-CFG representations at the RHS of (a) and (b) share the **projection** of GRs and differs in their **configuration** — while (a) generates an SVO order, (b) generates a so-called Verb-Initial (VI) construction.

(c)

S

VP-*PRD* ADVP NP-*SBJ* NP-*OBJ* DOT

S

{*PRD,OBJ,SBJ*}@S

PRD@S *PRD:SBJ*@S *SBJ*@S *OBJ*@S *OBJ:*@S
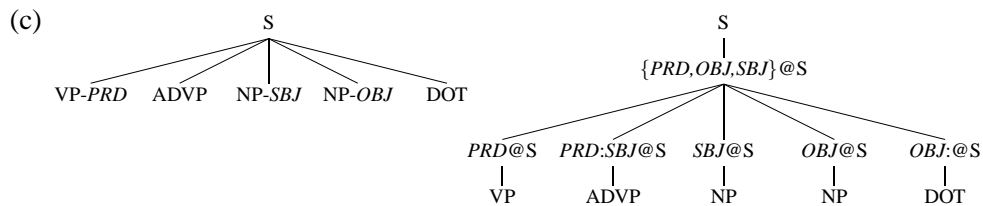
VP ADVP NP NP DOT

Figure 2: **Generating Adjunction and Periphrastic Configurations:** The CF depiction of S at the LHS of (c) generates complements, adjuncts, and punctuation in one go, whereas the RR-CFG representation at the RHS generates first the projection of core grammatical elements and then the configuration of a modified affirmative sentence in which they are realized. (Similarly, realising a question configuration using inversion in, e.g., English, naturally follows).

(d)

S

VP-*PRD* NP$_{[Def+,Acc+]}$-*OBJ* NP$_{[Def+]}$-*SBJ*

VB AT$_{[Acc+]}$ NP$_{[Def+]}$ NN$_{[Def+]}$

NNT NN$_{[Def+]}$

S

{*PRD,OBJ,SBJ*}@S

PRD@S *OBJ*@S *SBJ*@S

VP NP$_{[Def+,Acc+]}$ NP$_{[Def+]}$

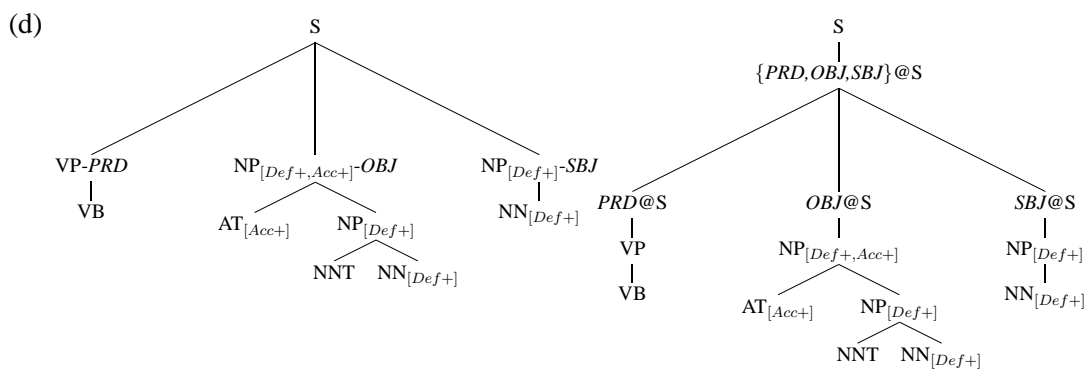VB AT$_{[Acc+]}$ NP$_{[Def+]}$ NN$_{[Def+]}$

NNT NN$_{[Def+]}$

Figure 3: **Realizing Grammatical Relations with bounded and unbounded Morphemes:** The CF depiction of the S level constituent at the LHS of (d) shows a strong dependence between the position of syntactic constituents and the morphologically realized features percolated from lower surface forms. In the *RR*-CFG representation at the RHS the feature distribution among sub constituents is dependent on grammatical relations, independently of their positioning. The realization stage generates a morphosyntactic paradigm in one go, allowing to capture meaningful collocations and idiosyncrasies, e.g., the Xor relation of the *Acc+* and *def+* features when marking direct objects in MH (Danon, 2001).
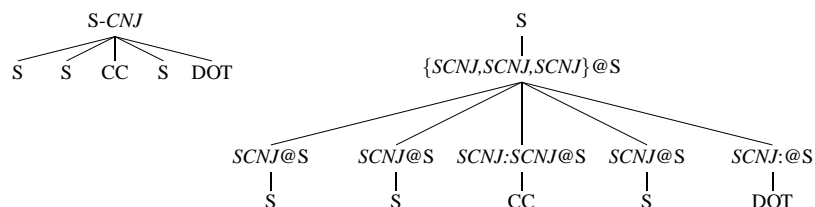
S-*CNJ*

S S CC S DOT

S

{*SCNJ,SCNJ,SCNJ*}@S

SCNJ@S *SCNJ*@S *SCNJ:SCNJ*@S *SCNJ*@S *SCNJ:*@S

S S CC S DOT

Figure 4: **Generating a Conjunction Structure:** The conjunction structure in the LHS of (e) is generated by the *RR*-CFG on the RHS in three stages. First, a relational network of finite number of conjuncts is generated, then a configuration for the conjuncts and conjunction markers (in MH, a CC before the last conjunct) is proposed, and finally the different conjuncts are generated conditioned on the same grammatical relation and the same parent. (Note that the possibility of different means for realizing conjunction, e.g., using morphemes, punctuation or multiple adpositions, falls out naturally from this setup.)

5 Experiments

Data The data we use is taken from the Modern Hebrew Treebank (MHTB) (Sima'an et al., 2001) which consists of 6501 sentences from the newspaper 'haaretz' annotated with phrase-structure trees and decorated with various morphological and functional features. We use version 2.0 of the treebank[3] which we processed and head-annotated as in (Tsarfaty and Sima'an, 2007). We experimented with sentences 1–500 (development set) and sentences 501–6001 (training set), and used sentences 6001-6501 (test set) for confirming our best result.

Models Our *Plain* models use the coarse-level MHTB category-labels enriched with various morphological features. Our morphological representation **Base** varies with respect to the use of the percolated features definiteness **Def** and accusativity **Acc**. Constituents' morphosyntactic representations enriched with their head PoS tag are referred to as *Head* and grand-parent encodings as *Parent*. For each combination of morphological and syntactic features we experimented with a state-split PCFG and with our *RR*-PCFG implementation.

Procedure We read off our models' parameters from the decorated phrase-structure trees in the MHTB, and use relative frequency estimation to instantiate their probability distributions. We smooth lexical rules using a PoS-tags distribution we learn for rare-words, where the 'rare' threshold is set to 1. We then use BitPar, a general purpose chart parser,[4] to find the most likely structures, and we extract the corresponding coarse-grained tree-skeletons for the purpose of evaluation.[5] We use PARSEVAL measures to quantitatively evaluate our models and perform a qualitative analysis of the resulting parse trees.

Results Table 1 shows the average F-Measure value for all sentences of length ≤40 in our development set with/without punctuation. The naïve baseline implementation for our experiments, the **Base***Plain* PCFG, performs at the level of 67.61/68.67 (comparable to the baseline reported in (Tsarfaty and Sima'an, 2007)). For all

Morphology	Syntax	Plain	Head	Parent	ParentHead
Base	(PCFG)	67.61/68.77	71.01/72.48	73.56/73.79	73.44/73.61
	(*RR*-PCFG)	65.86/66.86	71.84/72.76	74.06/74.28	75.13/75.29
BaseDef	(PCFG)	67.68/68.86	71.17/72.47	<u>74.13/74.39</u>	72.54/72.79
	(*RR*-PCFG)	66.65/67.86	73.09/74.13	74.59/74.59	76.05/76.34
BaseDefAcc	(PCFG)	68.11/69.30	71.50/72.75	74.16/ 74.41	72.77/73.01
	(*RR*-PCFG)	67.13/68.01	73.63/74.69	74.65/74.79	**76.15/ 76.43**

Table 1: **Parsing Results for Sentences of Length** < 40 **in the Development Set:** Averaged F-Measure With/Without Punctuation. **Base** refers to coarse syntactic categories, **Def** indicates percolating definiteness values, **Acc** indicated percolating accusativity marking. The underlined results replicate previously reported results in similar settings.

models in the *Plain* column the simple PCFG outperforms the *RR*-variety. Yet, the contribution of percolated morphological features is higher with the *RR*-PCFG than with the simple PCFG.

Moving to the *Head* column, we see that all *RR*-models already outperform their enriched PCFG counterparts. Again, morphological information contributes more to the *RR*-variety. The best result for this column, achieved by the **BaseDefAcc***Head* *RR*-model (63.73/64.69), outperforms its PCFG counterpart as well as all two-dimensional models reported by (Tsarfaty and Sima'an, 2007). In the *Parent* column), our *RR*-variety continues to outperform the PCFG albeit in an insignificant rate. (Both results are at the same level as the best model of (Tsarfaty and Sima'an, 2007).)

Finally, for all models in the *ParentHead* column the *RR*-models outperform their PCFG counterparts to a significant degree. Similarly to the *Head* column, the more morphological information is added, the greater the improvement is. Our best *RR*-model, **BaseDefAcc***ParentHead*, scores almost 10pt (25% error reduction) more than the *Plain* PCFG, it is about 3.5pt better (13% error reduction) than a state-split PCFG using the same information, and almost 2pt (7% error reduction) more than the best results reported for MH so far.

We confirmed the results of our best model on our test set, for which our baseline (**Base***Plain*) obtained 69.63/70.31. The enriched PCFG of **DaseDefAcc***HeadParent* yields 73.66/73.86 whereas the *RR*-PCFG yields 75.83/75.89. The overall performance for PCFGs is higher on this set, yet the *RR*-model shows a notable improvement (about 9% error reduction).

6 Analysis and Discussion

The trends in our quantitative analysis suggest that the *RR*-models are more powerful in exploiting different sorts of information encoded in parse

(a)

S — NP [CDT EFRWT "tens-of" / NP [NN ANFIM "people"]] — VP [VB MGIEIM "arrive"] — PP [IN M "from" / NP [NNP TAILND "Thailand"]] — PP [IN L "to" / NP [NNP IFRAL "Israel"]] ...

(b)

S — NP [CDT EFRWT "tens-of" / NP [NN ANFIM "people"]] — VP [VB MGIEIM "arrive"] — PP [IN M "from" / NP [NP [NNP TAILND "Thailand"] / PP [IN L "to" / NP [NNP IFRAL "Israel"]]]] ....
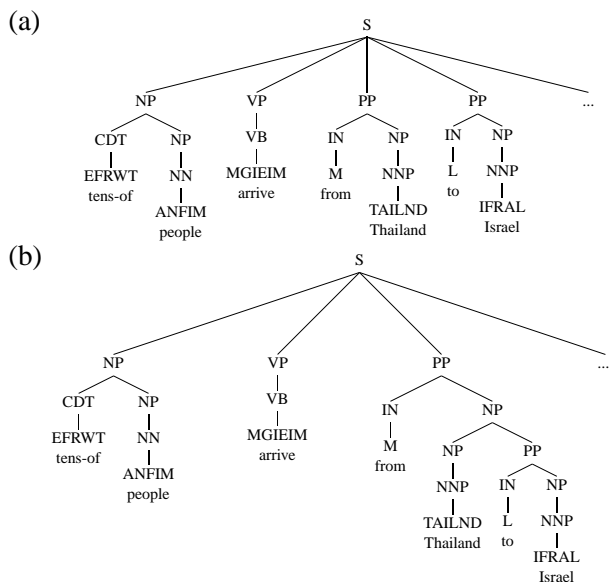
Figure 5: **Qualitative Analysis of Sentence (Fragment) #1:** (a) is the gold tree fragment, correctly predicted by our best *RR*-PCFG model. (b) is the tree fragment predicted by the PCFG corresponding to previously reported results.

(a)

S — PP [MCD FNI "on the other hand"] — VP [VB MTIR "allows"] — NP [NNP "the ministry of..."] — VP [VB LHESIK to-employ / NP [EWBDIM ZRIM "foreign workers"] / PP [B. in..]]

(b)

S — PP [IN M "from" / NP [NNT CD "side" / NP [CDT FNI "two" / NP [NNT MTIR "allows" / NP [NNP "the ministry of..."]]]]] — VP [LHESIK to-employ] — NP [NP [EWBDIM "workers"] / ADJP [ZRIM "foreigners"] / PP [B. in..]]
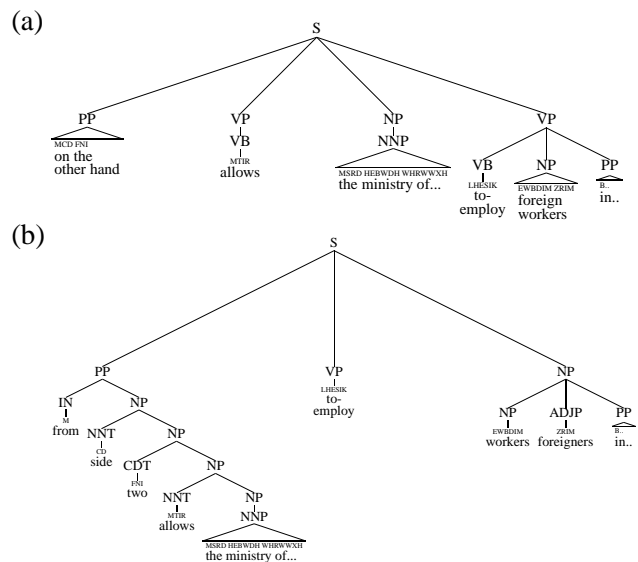
Figure 6: **Qualitative Analysis of Sentence (Fragment) #4:** (a) is the gold tree fragment, correctly predicted by our best *RR*-PCFG model. (b) is the tree fragment predicted by the PCFG corresponding to previously reported results.

trees, be it morphological information coming from dominated surface forms or functional information on top of syntactic categories.

We have shown that head information, which has very little contribution to parsing accuracy as a mere state-split, turns out to have crucial effects within the *RR*-models. For state-splits based PCFGs, adding head information brings about a category fragmentation and decreasing performance. The separation between form and function we articulate in the *RR*-approach allows us to capture generalizations concerning the distribution of syntactic constituents under heads based on their grammatical function, and use fine-grained features to predict their morphosyntactic behaviour.

We have further shown that morphological information contributes a substantial improvement when adopting the *RR*-approach, which is inline with the linguistic insight that there is a correlation between morphological marking on top of surface forms and the grammatical function their dominating constituents realize. Morphological information is particularly useful in the presence of heads. Taken together, head and percolated features implement a rather complete conceptualization of multiple exponence.

To wrap up the discussion, we leave numbers aside and concentrate on the *kind* of structures predicted by our best model in comparison to the ones suggested by previously reported unlexicalized PCFGs ((Tsarfaty and Sima'an, 2007), underlined in our table). Due to lack of space we

only discuss errors found within the first 10 parsed sentence, yet we note that the qualitative trend we describe here persists throughout our development set. Figures (5) and (6) show a gold tree (a fragment of sentence #1) correctly predicted by our best *RR*-model (a) in comparison with the one predicted by the respective PCFG (b). The tree fragment in figure (5) shows that the *RR*-grammar bracketed and attached correctly all the constituents that bear grammatical relations to the S clause (5a). The corresponding PCFG conflated the "to" and "from" phrases to a rather meaningless prepositional phrase (5b). For (a fragment of) sentence #4 in our set (figure 6) the *RR*-model recovered all grammatically meaningful constituents under the S clause and under the internal VP (6a). Notably, the PCFG in (6b) recovered *none* of them. Both grammars make attachment mistakes internal to complex NPs, but the *RR*-model is better at identifying higher level constituents that correlate with meaningful grammatical functions.

Our qualitative analysis suggests that our model is even more powerful than our quantitative analysis indicates, yet we leave the discussion of better ways to quantify this for future research.

A Note on Related Work Studies on parsing MH to date concentrate mostly on spelling out the integration of a PCFG parser with a morphological disambiguation component (e.g., (Tsarfaty, 2006; Goldberg and Tsarfaty, 2008)). On a setup identical to ours (gold segmentation, no PoS) the latter obtained 70pt. (Tsarfaty and Sima'an,

2007) examined the contribution of horizontal and vertical conditioning to an unlexicalized MH parser and concluded that head-driven Markovization performs below the level of vertical conditioning enriched with percolated features. We do not know of existing dependency-parsers applied to parsing MH or mildly-context-sensitive broad-coverage parsers applied to parsing a Semitic language.[6] To the best of our knowledge, this is the first fully generative probabilistic framework that models explicitly morpho-syntactic interaction to enhance parsing for non-configrational languages.

7 Conclusion

Projection and *Realization* are two sides of the same coin. *Projection* determines *which* grammatical relations appear in the syntactic representation, and *Realization* determines *how* such relations are realized. We suggest that the *Relational-Realizational* (*RR*) approach is adequate for parsing languages characteristically different from English, and we illustrate it with an application to parsing MH. We show that our approach to modeling the interaction between syntactic categories and a handful of percolated features already yields a notable improvement in parsing accuracy and substantially improves the quality of suggested parses. Incorporating additional functional and morphological information, we expect, will help bridging the gap in performance between Hebrew and configurational languages such as English.

References

Anderson, S. R. 1992. *A-Morphus Morphology*. Cambridge University Press.

Beard, R. 1988. The Separation of Derivation and Affixation: Toward a Lexeme-Morpheme Base Morphology. *Quarderni di semantica*, pages 277–287.

Blevins, J. P. 2008. Periphrasis as Syntactic Exponence. In Ackerman, F., J.P. Blevins, and G.S. Stump, editors, *Patterns in Paradigms*. CSLI.

Bod, R. 2003. An Efficient Implementation of a New Dop Model. In *Proceedings of EACL*.

Bresnan, Joan. 1979. *A Theory of Grammatical Representation*. Duplicated Lecture Notes. Department of Linguistics and Philosophy, MIT.

Charniak, E. 1997. Statistical Parsing with a Context-Free Grammar and Word Statistics. In *AAAI/IAAI*.

Collins, M. 2003. Head-Driven Statistical Models for Natural Language Parsing. *Comp. Linguistics*.

Danon, G. 2001. Syntactic Definiteness in the Grammar of Modern Hebrew. *Linguistics*, 6(39).

Goldberg, Y. and R. Tsarfaty. 2008. A Single Generative Framework for Joint Morphological Segmentation and Syntactic Parsing. In *Proceedings of ACL*.

Johnson, M. 1998. PCFG Models of Linguistic Tree Representations. *Computational Linguistics*, 24(4).

Klein, D. and C. Manning. 2003. Accurate Unlexicalized Parsing. In *Proceedings of ACL*.

McDonald, R., F. Pereira, K. Ribarov, and J. Hajič. 2005. Non-Projective Dependency Parsing using Spanning Tree Algorithms. In *Proceedings of HLT*.

Nivre, J. 2007. Data-driven Dependency Parsing Across Languages and Domains; Perspectives from the CoNLL-2007 Shared Task. In *Proceedings of IWPT*.

Perlmutter, D. M. 1982. Syntactic Representation, Syntactic levels, and the Notion of Subject. In Jacobson, Pauline and Geoffrey Pullum, editors, *The Nature of Syntactic Representation*. Springer.

Postal, P. M. and D. M. Perlmutter. 1977. Toward a Universal Characterization of Passivization. In *BLS3*.

Postal, P. M. 1982. Some Arc-Pair Grammar Decriptions. In Jacobson, P. and G. K. Pullum, editors, *The Nature of Syntactic Representation*. Dordrecht.

Sima'an, K., A. Itai, Y. Winter, A. Altman, and N. Nativ. 2001. Building a Tree-Bank for Modern Hebrew Text. In *Traitement Automatique des Langues*.

Tsarfaty, R. and K. Sima'an. 2007. Three-Dimensional Parametrization for Parsing Morphologically Rich Languages. In *Proceedings of IWPT*.

Tsarfaty, R. 2006. Integrated Morphological and Syntactic Disambiguation for Modern Hebrew. In *Proceeding of ACL-SRW*.

Zwicky, A. M. 1993. Heads, Bases, and Functors. In Corbett, G.G., N. Fraser, and S. McGlashan, editors, *Heads in Grammatical Theory*. Cambridge.

---

[6]Parsing MSA has been explored with a treebank three times as large as ours using a head-driven lexicalized parser obtaining around 78% accuracy (`http://papers.ldc.upenn.edu/`). The input setup assumes gold segmentation as well as PoS tags information and some diacritization.