GF	*Description*	*Applicable to...*
PRD	Predicative Elements	VP, PREDP
SBJ	Grammatical Subjects	NP, SBAR
OBJ	Direct Objects	NP
COM	Indirect Objects	NP, PP
	FInite Complements	SBAR
IC	Infinitival Complements	VP
CNJ	A Conjunct within a Conjunction Structure	All

Table 2: Grammatical Functions in the MHTB

SP-PCFG	Expansion	$P(C_{l_n}, \ldots, C_h, \ldots, C_{r_n} \mid P)$
HD-PCFG	Head	$P(C_h \mid P)$
	Left Branch?	$P(\mathbf{L}{:}\Delta_{l_1}, \mathbf{H}{:}\Delta_{r_1} \mid C_h, P)$
	Right Branch?	$P(\mathbf{R}{:}\Delta_{r_1} \mid \Delta_{l_n}, C_h, P)$
	Left Arg/Mod	$P(C_{l_i}, \Delta_{l_{i+1}} \mid \mathbf{L}, \Delta_{l_i}, C_h, P)$
	Right Arg/Mod	$P(C_{r_i}, \Delta_{r_{i+1}} \mid \mathbf{R}, \Delta_{r_i}, C_h, P)$
	Left Final?	$P(C_{l_1} \mid \mathbf{L}, \Delta_{l_{n-1}}, C_h, P)$
	Right Final?	$P(C_{r_n} \mid \mathbf{R}, \Delta_{r_{n-1}}, C_h, P)$
RR-PCFG	Projection	$P(\{gr_1, \ldots, gr_m\} \mid P)$
	Configuration	$P(\langle gr_1, \ldots, gr_m\rangle \mid \{gr_1, \ldots, gr_m\} P)$
	Realization	$P(C_{m_j} \mid gr_j, P)$
	Adjunction	$P(C_{j_1}, \ldots, C_{j_n} \mid gr_j, gr_{j+1}, P)$

Table 3: PCFG Parameter Classes for All Models

structure trees. In our version of the MHTB, *definiteness* and *accusativity* features are percolated from the PoS-tags level to phrase-level categories, extending the procedure of (Guthmann et al., 2009). For all models, we applied non-terminal state-splits distinguishing finite from non-finite verb forms and possessive from non-possessive noun phrases. We head-annotated the treebank, and based on the 'subject', 'object', 'complement' and 'conjunction' labels in the MHTB we devised an automatic procedure to annotate all the grammatical functions indicated in table 2.[7]

Procedure For all models, we learn a PCFG by reading off the parameters described in table 3, in accordance with the trees depicted in figures 1–3.[8] For all models, we use relative frequency estimates. For lexical parameters, we use a simple smoothing procedure assigning probability to unknown words using the per-tag distribution of rare words ("rare" threshold set to < 2). The input to our parser consists of morphologically segmented surface forms, and the parser has to assign the syntactic as well as morphological analysis to the surface segments.[9] We use the standard development/training/test split as in (Tsarfaty and Sima'an, 2008). Since our goal is a detailed comparison and fine-grained analysis of the results we concentrate on the development set. We use a general-purpose CKY parser (Schmid, 2004) to exhaustively parse the sentences, and we strip off all model-specific information prior to evaluation.

Evaluation We use standard *Parseval* measures calculated for the original, flat, canonical representation of the parse trees.[10] We report *Precision/Recall* for the coarse-grained non-terminal categories. In addition to overall Parseval scores we report the accuracy results *Per Syntactic Category*. We further report model size in terms of the number of parameters. As is well known in Machine Learning, models with more parameters require more data to learn, and are more vulnerable to sparseness. In our evaluation we thus follow the rule of thumb that (all else being equal) for models of equal size the better performing model is preferred, and for models with equal performance, the smaller one is preferred.

5 Results and Analysis

5.1 Overall Results

Table 4 shows the parsing results for the **State-Split (SP) PCFG**, the **Head-Driven (HD) PCFG** and the **Relational-Realizational (RR) PCFG** models on parsing the Modern Hebrew Treebank, with *definiteness* and *accusativity* marked on PoS-tags as well as phrase-level categories. For all models, we experiment with grandparent encoding. For non-HD models, we also examine the utility of a head-category split.[11]

[7] The enhanced corpus will be available at `www.science.uva.nl/~rtsarfat/resources.htm`.

[8] Our training procedure is strictly equivalent to the transform-detransform methodology of (Johnson, 1998), but we implement a tree-traverse procedure as in (Bikel, 2002) collecting all parameters per event at once.

[9] This setup is more difficult than, e.g., the Arabic parsing setup of (Bikel, 2002), as they assume gold-standard pos-tags as input. Yet it is easier than the setup of (Tsarfaty, 2006; Goldberg and Tsarfaty, 2008) which uses unsegmented surface forms as input. The decision to use segmented and untagged forms was made to retain a realistic scenario. Morphological analysis is known to be ambiguous, and we do not assume that morphological features are known up front. Morphological segmentation is also ambiguous, but for our purposes it is unavoidable. When comparing different models on an individual sentence they may propose segmentation to sequences of different lengths, for which accuracy results cannot be faithfully compared. See (Tsarfaty, 2006) for discussion.

[10] The flat canonical representation also allows for a fair comparison that is not biased by the differing branching factors of the different models.

[11] In HD models, a head-tag is already assumed in the conditioning context for sister nodes (Klein and Manning, 2003).