Model	SP-PCFG	HD-PCFG	HD-PCFG	HD-PCFG	HD-PCFG	RR-PCFG
Type Of Distance / Δ or Subcategorization	Phrase-Level State Splits	Intervening Verb/Punc	Left and Right #Constituents	Left and Right Constituent Labels	Left and Right Function Labels	Subcat Sets Configuration
Precision/Recall (#Params)	70.95/70.32 (13884)	72.39/71.97 (11650)	72.70/74.46 (18058)	72.42/74.26 (16534)	72.84/74.62 (16460)	76.32/76.51 (13618)

Table 6: **Incorporating Distance and Grammatical Functions into Head-Driven Parsing Models** Reporting Precison/Recall (#Parameters) for Sentences Length < 40.

6.1 Results and Analysis

Table 6 reports the results of experimenting with HD models with different instantiations of a *distance* function, starting from the standard notion of (Collins, 2003) and ending with our proposed, relational, function sets. For all HD models, we retain the *head, left* and *right* generation cycle and only change the conditioning context (Δ_i) for sister generation.

As a baseline, we show the results of adding grammatical function information as state-splits on top of an SP-PCFG.[13] This SP model presents much lower performance than the RR model although they are almost of the same size and they start off with the same information. This result shows that sophisticated modeling can blunt the claws of the sparseness problem. One may obtain the same number of parameters for two different models, but correlate them with more profound linguistic notions in one model than in the other. In our case, there is more statistical evidence in the data for, e.g., case marking patterns, than for association of grammatical relations with structurally-marked positions.

For all HD variations, the RR model continues to outperform HD models. The function-set variation performs slightly (but not significantly) better than the category-set. What seems to be still standing in the way of getting useful disambiguation cues for HD models is the fact that the *left* and *right* direction of realization is hardwired in their representation. This breaks down a coherent distribution over morphosyntactic representations realizing grammatical relations to arbitrary position-dependent fragments, which results in larger grammars and inferior performance.[14]

[13]The startegy of adding grammatical functions as state-splits is used in, e.g., German (Rafferty and Manning, 2008).

[14]Due to the difference in the size of the grammars, one could argue that smoothing will bridge the gap between the HD and RR modeling strategies. However, the better size/accuracy trade-off shown here for RR models suggests that they provide a good bias/variance balancing point, especially for feature-rich models characterizing morphologically rich languages. A promising strategy then would be to smooth or split-and-merge (Petrov et al., 2006)) RR-based models rather than to add an elaborate smoothing component to configurationally-based HD models.

7 A Typological Detour

Hebrew, Arabic and other Semitic Languages are known to be substantially different from English in that English is strongly *configurational*. In configurational languages word-order is fixed, and information about the grammatical functions of constituents (e.g., *subject* or *object*) is often correlated with structurally-marked positions inside highly-nested constituency structures. *Nonconfigurational* languages (Hale, 1983), in contrast, allow for freedom in their word-ordering and information about grammatical relations between constituents is often marked by means of *morphology*.

Configurationality is hardly a clear-cut notion. The difference in the configurationality level of different languages is often conceived as depicted in figure 7. In *linguistic typology*, the branch of linguistics that studies the differences between languages (Song, 2001), the division of labor between linear ordering and morphological marking in the realization of grammatical relations is often viewed as a continuum. Common wisdom has it that the lower a language is on the configurationality scale, the more morphological marking we expect to be used (Bresnan, 2001, page 6).

For a statistical parser to cope with nonconfigurational phenomena as observed in, for instance, Hebrew or German, it should allow for flexibility in the *form* of realization of the grammatical *functions* within the phrase-structure representation of trees. Recent morphological theories employ *Form-Function* separation as a widely-accepted practice for enhancing the adequacy of models describing variability in the realization of grammatical *properties*. Our results suggest that the adequacy of syntactic processing models is related to such typological insights as well, and is enhanced by adopting a similar form-function separation for expressing grammatical *relations*.