

values between *ssmin* and *ssmax*. The identifier of the other generic entity-type *text* is defined as *text.id* with a subset of the attributes of *all.unit.id* inherited in the 'isA' relationship between *node* and *text*. The attributes of the entity *XRef* are the keys of the citing and cited text units *citing.text.id* and *cited.text.id* respectively. These attributes draw their values from the domain *text.id*. Generic symbolic keys have been used throughout our work to provide a powerful mechanism for flexibly manipulating the complex object structure of figure 3 [Rossiter and Heather 1988].

3.4 Models for Expressing Dynamic Aspects

Major deficiencies of the E-R and Borkin models are that they have no defined operations and thus cannot handle dynamic control of the life-cycle of entities. Semantic models which enable such dynamic structures to be expressed as well as static ones have therefore also been examined such as Taxis, the Event Model and SHM+. The model Taxis illustrates the potential of this approach and its use is being explored at Newcastle for control of the legal drafting process [Rossiter and Heather 1990].

4 Basic Models and Text Structures

4.1 The Relational Model

None of the basic models is rich enough in capability to satisfy all requirements in manipulating complex textual structures. Hierarchical and network models can be quickly discounted but a more detailed discussion is necessary to illustrate weaknesses in a relational approach 'flattening' the data. The relational model offers the power of the network model but with a simple and elegant method of data manipulation. The E-R model given earlier for law can be implemented directly by mapping the table-types in figure 2(b) on to conventional relations: one table-type per relation. There is a major difficulty in this transformation for text structures: the unnormalized data for figure 2(b) cannot be retained in the relational model. Also, the semantics of the E-R diagram are now represented implicitly; for instance, the structure of the objects:

act -> schedule -> paragraph -> subparagraph -> word.placement

is represented by a series of relations whose attributes carry the inter-object relationships so that the basic hierarchical structure is not explicitly conveyed to the user as in the E-R diagram. For users with detailed knowledge