Model 1: Joint Relational Model (JRM)

The JRM uses a large corpus to learn the selectional preferences of a binary semantic relation by considering its arguments jointly.

Given a relation p and large corpus of English text, we first find all occurrences of relation p in the corpus. For every instance <*x, p, y*> in the corpus, we obtain the sets $C(x)$ and $C(y)$ of the semantic classes that x and y belong to. We then accumulate the frequencies of the triples <$c(x)$, p, $c(y)$> by assuming that every $c(x) \in C(x)$ can co-occur with every $c(y) \in C(y)$ and vice versa. Every triple <$c(x)$, p, $c(y)$> obtained in this manner is a candidate selectional preference for p. Following Pantel et al. (2007), we rank these candidates using Pointwise mutual information (Cover and Thomas 1991). The ranking function is defined as the strength of association between two semantic classes, c_x and c_y[2], given the relation p:

$$pmi(c_x|p; c_y|p) = \log \frac{P(c_x, c_y|p)}{P(c_x|p)P(c_y|p)} \tag{3.1}$$

Let $|c_x, p, c_y|$ denote the frequency of observing the instance <$c(x)$, p, $c(y)$>. We estimate the probabilities of Equation 3.1 using maximum likelihood estimates over our corpus:

$$P(c_x|p) = \frac{|c_x, p, *|}{|*, p, *|} \quad P(c_y|p) = \frac{|*, p, c_y|}{|*, p, *|} \qquad P(c_x, c_y|p) = \frac{|c_x, p, c_y|}{|*, p, *|} \tag{3.2}$$

We estimate the above frequencies using:

$$|c_x, p, *| = \sum_{w \in c_x} \frac{|w, p, *|}{|C(w)|} \quad |*, p, c_y| = \sum_{w \in c_y} \frac{|*, p, w|}{|C(w)|} \qquad |c_x, p, c_y| = \sum_{w_1 \in c_x, w_2 \in c_y} \frac{|w_1, p, w_2|}{|C(w_1)| \times |C(w_2)|} \tag{3.3}$$

where $|x, p, y|$ denotes the frequency of observing the instance <x, p, y> and $|C(w)|$ denotes the number of classes to which word w belongs. $|C(w)|$ distributes w's mass equally among all of its senses $C(w)$.

Model 2: Independent Relational Model (IRM)

Due to sparse data, the JRM is likely to miss some pair(s) of valid relational selectional preferences. Hence we use the IRM, which models the arguments of a binary semantic relation independently.

Similar to JRM, we find all instances of the form <*x, p, y*> for a relation p. We then find the sets $C(x)$ and $C(y)$ of the semantic classes that x and y belong to and accumulate the frequencies of the triples <$c(x)$, p, *> and <*, p, $c(y)$> where $c(x) \in C(x)$ and $c(y) \in C(y)$.

All the tuples <$c(x)$, p, *> and <*, p, $c(y)$> are the independent candidate RSPs for a relation p and we rank them according to equation 3.3.

Once we have the independently learnt RSPs, we need to convert them into a joint representation for use by the inference plausibility and directionality model. To do this, we obtain the Cartesian product between the sets <$C(x)$, p, *> and <*, p, $C(y)$> for a relation p. The Cartesian product between two sets A and B is given by:

$$A \times B = \{(a,b) : \forall a \in A \quad and \quad \forall b \in B\} \tag{3.4}$$

Similarly we obtain:

$$\langle C_x, p, * \rangle \times \langle *, p, C_y \rangle = \left\{ \begin{matrix} \langle c_x, p, c_y \rangle : \ \forall \langle c_x, p, * \rangle \in \langle C_x, p, * \rangle \quad and \\ \forall \langle *, p, c_y \rangle \in \langle *, p, C_y \rangle \end{matrix} \right\} \tag{3.5}$$

The Cartesian product in equation 3.5 gives the joint representation of the RSPs of the relation p learnt using IRM. In the joint representation, the IRM RSPs have the form <$c(x)$, p, $c(y)$> which is the same form as the JRM RSPs.

3.3 Inference plausibility and directionality model

Our model for determining inference plausibility and directionality is based on the intuition that for an inference to hold between two semantic relations there must exist sufficient overlap between their contexts and the directionality of the inference depends on the quantitative comparison between their contexts.

Here we model the context of a relation by the selectional preferences of that relation. We determine the plausibility of an inference based on the overlap coefficient (Manning and Schütze, 1999) between the selectional preferences of the two paths. We determine the directionality based on the difference in the number of selectional preferences of the relations when the inference seems plausible.

Given a candidate inference rule $p_i \Leftrightarrow p_j$, we first obtain the RSPs <$C(x)$, p_i, $C(y)$> for p_i and <$C(x)$, p_j, $C(y)$> for p_j. We then calculate the overlap coefficient between their respective RSPs. Overlap coefficient is one of the many distribu-

[2] c_x and c_y are shorthand for $c(x)$ and $c(y)$ in our equations.