

- Converted nouns: nouns which show the identical phonological form and word shape as the corresponding verbs with the exception that their initial letter is capitalized, e.g. <Essen> [meal] vs. <essen> [to eat]. The nominal word category for this type of word can only be clearly identified on the basis of the syntactic structure of the sentence in which the respective noun occurs.

The analysis of the structure of German nouns can be summarized as follows: Two groups of nouns can be distinguished with one group containing simple nouns and derived nouns. Their nominal word category can be identified on the word level, either because the word category information is inherent in the noun itself, as is the case for simple nouns, or because the word contains a (nominal) suffix, as is the case for derived nouns. The second group contains words (converted nouns) which, seen or heard in isolation, cannot be unambiguously identified as nouns. This group of words is ambiguous in respect of its word category as they can be used as verbs as well as nouns. According to ordered access models, i.e. frequency-dependent models (e.g., Foster & Bednall, 1976), it can be assumed that, when an ambiguous word is presented in the absence of context, the more frequent word category alternative is retrieved first. Evidence for frequency-dependent lexical retrieval comes from studies of homophones/homographs. It was found that the more frequent meaning of a homophone/homograph was retrieved quicker than its less frequent meaning. (Simpson & Burgess, 1985; for a review, see Hillert, 1997). Examples of frequency distributions for this type of words in German are given in table 1.

	noun/verb	noun, absolute frequency	verb, absolute frequency	noun, relative, frequency
MISSTRAUEN	mistrust/mistrust	107	3	97,3%
LEBEN	life/live	2558	700	78,5%
VERSPRECHEN	promise/promise	69	58	54,3%
SCHREIBEN	writing/write	195	350	35,8%
BEMÜHEN	endeavour/endeavour	0	146	0%

Table 1: Frequency of words used as verbs as well as nouns in written language, CELEX

Therefore, the question posed is, how are these ambiguous words processed when they are presented in a sentence context. Are both alternative word category entries retrieved with no regard for syntactical information or is the syntactic context used from the outset to retrieve only the appropriate word category? Previous studies into the lexical ambiguity of

No. Veh.	1	2	1 - 3	1 - 4	1 - 5	1 - 6	Av. Raw Matches	σ Raw Matches	Av. Cor. Matches	σ Cor. Matches
1000	10						111.4	8.5	11.4	8.5
2000	10						411.8	19.5	11.8	19.5
1000	100						199.2	12.0	99.2	12.0
1000	200						302.3	7.7	202.3	7.7
1000	500						596.6	12.3	496.7	12.3
1000	0	10					21.9	4.6	9.3	3.3
1000	500	10					73.8	7.5	10.2	6.2
1000	100	100					152.1	8.5	101.9	7.5
1000	500	250					388.3	22.7	253.2	20.1
1000	0	500					667.2	24.9	506.0	22.3
1000	0	0	100				154.6	26.6	104.0	22.6
1000	100	100	100				164.4	11.4	97.7	9.3
500	100	100	100	100			140.7	19.3	105.8	17.4
1000	500	250	100				207.8	29.7	106.1	23.7
500	10	10	10	10			14.2	2.2	10.5	1.8
1000	10	10	10	10			17.4	4.1	9.4	2.8
500	50	50	50	50			71.3	14.3	47.8	12.3
500	100	100	100	100			151.9	26.9	92.0	22.3
1000	0	0	0	100			177.6	29.9	103.4	22.6
1000	100	100	100	100			222.2	61.5	111.0	46.7
1000	0	0	0	0	10		21.2	13.4	12.3	9.9
500	0	0	0	0	100		152.6	45.5	92.2	37.3
1000	0	0	0	0	100		214.6	58.0	103.5	40.2
1000	100	100	100	100	100		289.8	88.4	101.3	55.0

Table 1: Simulation results — all performed over twenty runs with 10,000 distinct vehicle types.

problems with the standard deviation being higher than would be desirable in real cases. It is important to bear in mind that these were relatively extreme tests of the method since $p(2)$ and $p(3)$ were relatively low and the number of samples given were quite high. Often the method was attempting to predict only ten true matches in a number of observed matches which might be several hundred.

8 CONCLUSIONS

This paper presented a framework for analysis of surveys where matches are required over more than two data collection points. The framework given formalises the concept of a type of match using the concept of the equivalence class. Further a method is given for evaluating \mathcal{M}_n the set of all possible types of match over multiple data sets. An algorithm is given which shows how, computationally, \mathcal{M}_n can be computed and a method is given for enumerating its elements using Stirling numbers.

The framework given is then applied to the problem of false matches — which is put into the language of set theory using the concept of a partial ordering. It is shown how this partial ordering can be used to visualise, by means of a Hasse diagram, the ways in

Protein-ligand Complex	# rotatable bonds	AutoDock	DOCK	FlexX	ICM	GOLD	T10	T20
3ptb	3	0.80	0.59	1.11	0.49	1.09	0.56	0.54
1tng	2	0.62	0.86	1.08	0.71	1.89	0.70	0.69
1tnj	3	1.21	1.56	1.73	2.17	1.90	1.42	1.50
1tnk	4	1.69	1.87	1.70	2.53	3.08	1.16	1.14
1tni	5	2.61	5.26	2.73	3.40	4.93	2.22	2.22
1tpp	7	1.80	3.25	1.95	1.71	2.33	2.43	2.53
1pph	11	5.14	3.91	3.27	1.44	4.23	4.00	0.53
1phf	1	2.09	2.39	4.68	1.23	4.42	1.20	1.20
1phg	5	3.52	5.57	4.87	0.46	4.20	1.07	1.08
2cpp	3	3.40	2.48	0.44	2.53	3.49	3.26	3.27
1nsc	12	1.40	4.86	6.00	1.80	1.02	1.47	1.40
1nsd	11	1.20	4.51	1.56	1.04	0.96	1.85	1.85
1nnb	11	0.92	4.51	0.92	1.09	0.84	1.67	3.97
1cbx	5	1.33	3.13	1.32	0.82	1.87	0.62	0.62
3epa	8	2.22	6.48	1.51	0.77	1.87	2.22	2.22
6epa	16	8.30	8.30	9.83	1.60	4.96	4.00	4.00
1abe	4	0.16	1.87	0.55	0.36	0.18	0.56	0.56
1abf	5	0.48	3.25	0.76	0.61	0.50	0.68	0.70
5abp	6	0.48	3.89	4.68	0.88	0.59	0.48	0.51
1etr	15	4.61	6.66	7.26	0.87	5.99	1.09	1.09
1ets	13	5.06	3.93	2.11	6.22	2.39	1.97	1.97
1ett	11	8.12	1.33	6.24	0.99	1.30	0.82	0.82
3tmn	10	4.51	7.09	5.30	1.36	3.96	3.65	3.65
5tln	14	5.34	1.39	6.33	1.42	1.60	1.21	1.21
6tmn	20	8.72	7.78	4.51	2.60	8.54	2.21	2.21
1apt	30	1.89	8.06	5.95	0.88	8.82	5.72	4.79
1apu	29	9.10	7.58	8.43	2.02	10.70	1.32	1.32
2ifb	15	3.09	1.43	8.94	1.04	2.61	2.09	5.19
1eil	6	5.81	2.78	3.52	2.00	6.04	1.86	1.86
1okl	5	8.54	5.65	4.22	3.03	3.55	2.84	2.84
1cnx	13	10.9	7.35	6.83	2.09	6.32	6.20	6.20

Table 3. Comparison of best RMSD for different docking methods. The best RMSD is the RMSD of the predicted ligand from the Xray structure. For each protein-ligand complex, the best RMSD found is reported in bold.

4.4 Comparison of RMSD for the Different Docking Methods

The RMSD's reported in Figure 3 are the root mean square deviations of the heavy atoms of the predicted ligands from the corresponding ligands in their published complex crystal structures. For our MD-based docking, we present results of attempts with different numbers of trials: *T10* with 10 trials per attempt and *T20* with 20 trials per attempt. In general, we observe that for both *T10* and *T20*, we get, on the average, lower

RMSD than the other methods.

4.5 Comparison of Simulation Time for the Different Docking Methods

The main question we want to address in Table 4 is whether the high level of accuracy is also supported by competitive execution time when compared with the execution times of the other docking methods. Table 4 shows the average CPU time to complete a protein-ligand docking for the ten proteins in Table 1 and for