

Table 1: Examples of constructed nominal case frames.

	Case slot	Examples with freq	Generalized examples with rate
<i>kakaku</i> (1) (price)		Definition: the amount of money you have to pay for something .	
	[something]	<i>shôhin</i> (goods):9289, <i>seihin</i> (product):2520, <i>buhin</i> (part):341, <i>yunyuhin</i> (importation):232, ...	[CT:ARTIFACT]:0.93, ...
<i>yane</i> (1) (roof)		Definition: the structure that covers or forms the top of a building etc.	
	[building]	<i>ie</i> (house):2505, <i>kuruma</i> (car):1565, <i>koya</i> (hut):895, <i>tatemono</i> (building):883, <i>minka</i> (private house):679, ...	[CT:FACILITY]:0.44, [CT:VEHICLE]:0.13, ...
<i>shusho</i> (1) (prime minister)		Definition: the elected leader of the government in a country that has a parliament.	
	[country]	<i>nihon</i> (Japan):2355, <i>kuni</i> (country):272, <i>doitsu</i> (Germany):157, <i>chûgoku</i> (China):130, ...	[NE:LOCATION]:0.82, [CT:VEHICLE]:0.13, ...
<i>imouto</i> (1) (sister)		Definition: a girl or woman who has the same parents as you.	
	<relationship>	<i>watashi</i> (me):3385, <i>ore</i> (me):1188, <i>boku</i> (me):898, <i>jibun</i> (oneself):341, <i>tomodachi</i> (friend):537, ...	[CT:PERSON]:0.74, [NE:PERSON]:0.22, ...
<i>rebâ</i> (1) (lever)		Definition: a stick or handle on a machine .	
	[machine]	<i>buŕeki</i> (brake):122, <i>sokketo</i> (socket):67, <i>waipâ</i> (wiper):54, <i>souchi</i> (device):52, ...	[CT:ARTIFACT]:0.61, [CT:VEHICLE]:0.04, ...
<i>rebâ</i> (2) (liver)		Definition: the liver of an animal , used as food.	
	[animal]	<i>niwatori</i> (chicken):153, <i>buta</i> (pig):153, <i>ushi</i> (cattle):62, <i>doubutsu</i> (animal):25, ...	[CT:ANIMAL]:0.98, ...
<i>senshu</i> (1) (player)		Definition: someone who takes part in a sport .	
	[sport]	<i>yakyû</i> (baseball):1252, <i>rirê</i> (relay):736, <i>kyôgi</i> (competition):430, <i>sakkâ</i> (soccer):394, ...	[CT:ABSTRACTION]:0.56, ...
	<affiliation>	<i>chîmu</i> (team):4409, <i>nihon</i> (Japan):3222, <i>reddu</i> (Reds):771, <i>kankoku</i> (Korea):644, <i>rîgu</i> (league) ...	[NE:LOCATION]:0.33, [CT:ORGANIZATION]:0.30, ...

* “[] ” and “ < > ” denote dictionary-based and semantic feature-based analysis respectively. For details see (Sasano et al., 2004).

large corpora. Dictionary definition sentences are an informative resource to recognize obligatory cases of nouns. However, it is difficult to resolve associative anaphora by using a dictionary as it is, because all nouns in a definition sentence are not an obligatory case, and only the frequency information of noun phrases tells us which is the obligatory case. On the other hand, a simple method that just collects and clusters “ N_m no N_h ” phrases based on some similarity measure of nouns cannot construct comprehensive nominal case frames, because of polysemy and multiple obligatory cases. For details see (Sasano et al., 2004).

It is desirable to use a probability distribution for deciding whether a case slot is obligatory or not. However, it is difficult to estimate a probability distribution, since we construct nominal case frames not by using the examples of associative anaphora itself but by using the examples of noun phrases “ N_m no N_h ” (N_h of N_m). We use such noun phrases because indispensable entities of noun “ N_h ” often appear as “ N_m .” However, we can say neither frequently appeared “ N_m ” is an indispensable entity of “ N_h ,” nor an indispensable entity frequently appears as “ N_m .” For example, the name of a country is considered as an indispensable entity of “*shusho*” (prime minister), but

does not frequently appear as “ N_m .”² Thus, it is difficult to estimate a probability distribution and we use a hard decision.

2.2 NCF Construction from the Web

We constructed nominal case frames from the Web Corpus (Kawahara and Kurohashi, 2006), which comprises 1.6 billion unique Japanese sentences. In this corpus, there were about 390 million noun phrases “ N_m no N_h ,” about 100 million unique noun phrases, and about 17 million unique head nouns “ N_h .” There were about 4.07 million head nouns that appeared more than 10 times in the corpus, and we used only such head nouns.

The resultant nominal case frames consisted of about 564,000 nouns including compound nouns. We show examples of constructed nominal case frames in Table 1. The average number of case frames for a noun that has case frames was 1.0031, and the average number of case slots for a case frame was 1.0101. However, these statistics differed with the frequency of the noun. Therefore, we investigated the statistics of constructed nominal case frames for each group classified by the frequency of the nouns. Table 2 shows the re-

²It is because “the prime minister of Japan” is often mentioned by simply “the prime minister” in Japanese.