Chinese CogBank: Where to See the Cognitive Features of Chinese Words

Bin Li[*,†] **Xiaopeng Bai**[†] **Siqi Yin**[*] **Jie Xu**[*]
[*]School of Chinese Language and Literature
Nanjing Normal University
Nanjing, PR China
[†] Department of Computer Science
Brandeis University
MA, USA
libin.njnu@gmail.com, xpbai@brandeis.edu,
nnuyinsiqi@126.com, xujie.njnu@gmail.com

Abstract

Metaphor processing has been a heated topic in NLP. Cognitive properties of a word are important in metaphor understanding and generation. But data collected automatically tend to be reduced in both quantity and quality. This paper introduces CogBank a database of Chinese concepts and their associated cognitive properties. The database was constructed using simile templates to extract millions of "word-property" pairs via search engine over the World Wide Web. A method of manual check and correction was then implemented, resulting in the current CogBank database which contains 232,590 "word-property" pairs. CogBank also provides various search and visualization services for observing and comparing associations between concepts and properties.

1 Introduction

Metaphor studies in cognitive linguistics focus on the mechanisms of how metaphor works. Conceptual Metaphor Theory summarizes the types of mappings from source domain to target domain like "Time is Money" (Lakoff & Johnson, 1980). Blending Theory examines how the input spaces of two concepts blend new spaces (Fauconnier & Turner, 2003). Both theories emphasize the properties of a concept, which could be profiled in metaphor use. For example, *money* has properties of important, valuable and soulless that will help people to comprehend *time* in the metaphor. Many of these properties reflect common cognitive knowledge rather than scientific knowledge. If such cognitive properties can be collected and organized, it will benefit metaphor generation and understanding in NLP. However, manual construction of such databases could be time-consuming. In addition, the properties of concepts may vary from person to person. *Money* may have more than three properties and each property could be interpreted in different ways. This translates into three key issues to be solved: (1) How to collect as many concepts and properties as possible; (2) How to assure the properties are acceptable to native speakers; and, (3) How to evaluate the importance of the properties for a given concept.

Chinese CogBank is a database of cognitive properties of Chinese words. It has 232,590 "word-property" pairs, which consist of 82,937 words and 100,271 properties. The data were collected via Baidu.com, and adjudicated manually. Consequently, each "word-property" type has an associated frequency which can stand as a functional measure of the importance of a property.

The rest of the paper is organized as follows. Section 2 briefly reviews related work on collecting cognitive features. Section 3 introduces the construction of the Chinese CogBank. Descriptive statistics, search visualization tools of the database are presented in Sections 4 and 5. Section 6 discusses the potential applications of CogBank and the difficulties with respect to metaphor processing. Conclusions and future work are outlined in Section 7.

Proceedings of the Third Workshop on Metaphor in NLP, pages 77–86,
Denver, Colorado, June 5, 2015.

2 Related Work

Collecting the cognitive properties by hand can be tedious, time-consuming and problematic in terms of guaranteeing agreement between different annotators. Therefore corpus and web data have been taken as important resources. Kintsch (2000) collects noun-adjective pairs like "money-valuable" using Latent Semantic Analysis (LSA) from large corpora. Roncero et al. (2006) extracts noun-adjective pairs using the simile template "as adjective as noun". Using the same template, Veale & Hao (2007) collects English similes by querying Google using the nouns and adjectives in WordNet. Then the data contribute to a lexical metaphor knowledge base "sardonicus", which contains about 10,000 items of "noun-adjective" pairs. Veale et al. (2008) collects 25,000 items consisting of Chinese "noun-adjective" pairs from Google using words in HowNet. In a similar way, Jia & Yu (2009) collects Chinese similes from Baidu, yielding about 20,000 "noun-property" pairs.

At this stage, collection of "concept-property" pairs seems to reach a bottleneck in that it becomes difficult to substantially increase the number of valid items. The store of raw data collected is massive, ordinarily amounting to millions of items. Obviously, stores of data this massive contain much noise. Consulting each word in a normal dictionary would be a simple and efficient way to filter out noisy data. However the cost of such an approach would be that many good candidates are eliminated as they are not in the dictionary. Using a larger dictionary offers only limited improvement because many good candidates consist of multi-word expressions like "as sly as **jungle cat**", or even appear as embedded clauses such as "(Someone who is) as sly as **a fox is cunning and experienced**". Due to such difficulties, a large cognitive database is currently not available.

In addition, previous implementations have given little importance to word-property frequencies. It is important for a metaphor processing system to know how strong the relationship between the concept and the property. If the system must generate a metaphor expressing something that is white, it could find the most relevant concepts like snow and paper using collocation frequencies.

3 Construction of the Chinese CogBank

Like Roncero et al. (2006), Veale et al. (2008) and Jia & Yu (2009), we use simile templates to collect Chinese "word-property" items by querying the search engine Baidu. The lexicon items in HowNet are used to fill the simile templates.

3.1 Lexical Resources

HowNet is a structured Chinese-English bilingual lexical resource (Dong & Dong, 2006). Different from the synsets in WordNet (Miller, 1990), it describes a word by a set of structured semantic features named "sememe". About 2200 sememes are used to define 95,000 Chinese words and 85,000 English words in HowNet (ver. 2007). For example, the Chinese noun 猪(zhu) is translated to hog, pig and swine in English. The definition of 猪(zhu) is the sememe *livestock*|*牲畜*. A sememe is an English-Chinese combined label and is organized in a hierarchy. *livestock*|*牲畜 has its* hypernym sememe *animal*|*兽* and higher hypernym sememes *AnimalHuman*|*动物*, *animate*|*生物*, etc.

3.2 Data Collection

In Chinese, there are three simile templates which can be used to obtain the "word-property" pairs: "像 (as) + NOUN + 一样 (same)", "像 (as) + VERB + 一样 (same)" and "像 (as) + 一样 (same) + ADJ". We populated these with 51,020 nouns, 27,901 verbs and 12,252 adjectives from HowNet to query Baidu (www.baidu.com). Different from Veale et al. (2008), we included verbs because verbs as well as nouns have concept properties. For example, "抽筋(cramp)" is a verb in Chinese. It has the property "疼(painful)", which refers to people's experience in a cramp.

We submit 91,173 queries to Baidu, allowing up to 100 returned results for each query. Then 1,258,430 types (5,637,500 tokens) of "word-adjective" pairs are collected. Within such a large data set there will be many incoherent pairs. We filter out such pairs automatically via the nouns, verbs and adjectives in HowNet, resulting in a remaining 24,240 pairs. The words cover 6,022 words in HowNet, and the properties cover 3,539 words in HowNet. The high quality of these remaining pairs provides the potential for interesting

results. With the frequency information, we can see the top 10 most frequent pairs that fit the intuition of Chinese native speakers (see Table 1).

ID	Word	Property	Freq
1	苹果 apple	时尚 fashionable	1445
2	呼吸 breath	自然 natural	758
3	晨曦 sun rise	朝气蓬勃 spirited	750
4	纸 paper	薄 thin	660
5	雨点 rain drop	密集 dense	557
6	自由 freedom	美丽 beautiful	543
7	雪 snow	白 white	521
8	花儿 flower	美丽 beautiful	497
9	妖精 spirit	温柔 gentle	466
10	大海 sea	深 deep	402

Table 1. Top10 Most Frequent Word-Property Pairs

It might be surprising to see that "苹果 apple-时尚 fashionable" ranks top of all pairs. However, it makes sense because Apple (the brand) products are popular in China. "妖精 spirit" often refers to a young female demon/spirit who seduces people in Chinese fairy tales. The remaining 8 words represent ordinary things people experience in everyday life.

3.3 Manual Data Check

It is painful that in 5 million raw data only 24,240 pairs are left when filtered by HowNet. As stated in Section 2, we find a more productive way to increase the quantity of the database is to manually check the original data item by item.

To that end, we develop a set of guidelines for adjudication. We obtain four types of pairs from the sentences in the raw data. First, phrases like "as lazy as pig" contain good pairs, which we tagged as NORMAL. Second, pairs from phrases like "as valuable as ash" are tagged as IRONY. Third, pairs from sentences like "as soon as possible", "as fast as I can" are tagged as ELSE. The last type is ERROR in sentences like "as lazy as…".

After the manual correction, 843,086 pairs are left. As shown in Table 2, 232,590 are NORMAL items, 1,351 are IRONY. The rate of IRONY is much lower than the English data collected by Veale & Hao (2007). The reason is not clear yet. It may due to the different simile templates used in two languages. The other two categories ELSE and ERROR are uninformative for present purposes. But we find some important phenomena in the results that will be introduced in section 4.2.

Type	Num	Example
NORMAL	232,590	as lazy as pig
IRONY	1351	as valuable as ash
ELSE	389639	as soon as possible
ERROR	219506	as lazy as…
SUM	843,086	

Table 2. Four Kinds of Word-Property Pairs

4 Statistics

We find the results after adjudication to be better in both quality and quantity, generating 232,590 NORMAL pairs as the basis of the Chinese CogBank. In this section, we discuss the differences between the method of adjudication and automatic filtering of the data. We also present the descriptive statistics of CogBank.

4.1 Statistics of CogBank

Chinese CogBank has 232,590 "word-property" pairs, which consists of 82,937 words and 100,271 properties. The words cover 7,910 HowNet words, and the properties cover 4,376 HowNet words. This indicates that many more words and properties are gathered. Here we examine how much the results change compared to the filtered data in Section 3.2. Table 3 shows the top10 most frequent word-property pairs in CogBank. The result is not substantially different. The first item has changed to "freedom-beautiful", but "apple-fashionable" still ranks high in the database. Notably, the most frequent pairs are quite similar across automatic filtering and manual data check. In other words, if one only cares about the most frequent items from the web, automatic filtering is a fast and accurate method.

ID	Word	Property	Freq
1	自由 freedom	美丽 beautiful	3285
2	铁轨 rail track	长 long	2333
3	纸 paper	薄 thin	1828
4	天使 angel	美丽 beautiful	1766
5	苹果 apple	时尚 fashionable	1764
6	妖精 spirit	温柔 gentle	1565
7	阳光 sunlight	温暖 warm	1389
8	梦 dream	自由 free	1384
9	水晶 crystal	透明 clear	1336
10	雪 snow	白 white	1210

Table 3. Top10 Most Frequent Word-Property Pairs

Next we explore what the most frequent words and properties are in Chinese CogBank. This is important as we could learn what the most common entities are that people tend to use as vehicles in similes, and the most common properties people prefer to express in everyday life. As shown in Table 4, nouns like *flower, man, water, child, human, cat, angel, wolf* and *sunshine* rank the highest in the database. These words are quite common in everyday life and they have hundreds of properties. But the top 10 properties of each word dominate more than half the occurrences of these words when employed in a simile. This indicates that people always rely more heavily on a word's salient properties to form a simile expression.

Word	# of Pros	Freq	Top 10 Properties
花儿 flower	254	16991	绽放 bloom_7809,开放 bloom_1729,美丽 beautiful_1202, 红 red_965, 盛开 bloom_681, 美 beautiful/pretty_591, 灿烂 effulgent_561, 开 bloom_436, 香 sweet_278,简单 simple_220
花 flower	268	16602	绽放 bloom_6419,盛开 bloom_5375,美丽 beautiful_864, 美 beautiful/pretty_509, 开 bloom_435, 灿烂 effulgent_391, 开放 bloom_353, 多 numerous_148, 飘舞 dance in wind_125,漂亮 beautiful_92
男人 man	758	14708	战斗 fight_8771,奋斗 strive_975,拼命 desperate_234, 坚强 strong_213, 踢球 play football_130, 挑 pick_115, 活着 live_110, 打球 play ball_105, 裸上身 half naked_102,恋爱 in love_95
水 water	884	11837	流 flow_1786,流淌 flowing_697,流动 flow_524, 稀 dilute_380, 流过 flowing_323, 透明 limpid_245, 温柔 gentle_183, 清澈 limpid_176, 清淡 mild_170,泼 splash_168
孩子 child	1642	10866	快乐 happy_420, 哭 cry_352, 天真 childlike/innocent_332, 无助 helpless_233,说真话 tell the truth_229, 哭泣 cry/weep_216,好奇 curious_197, 兴奋 excited_172,笑 smile_167,开心 happy_166
人 human	1482	9468	活着 live_609,穿衣服 wear clothe_430, 生活 live_336,思考 think_316,直立行走 bipedalism/walk upright_315, 活 live_310, 说话 speak/talk_284, 走路 walk_222, 站立 stand_188, 站 stand_135
猫 cat	828	6989	蜷缩 curl_256,可爱 cute/lovely147,蹭 rub_137, 慵懒 lazy_136, 温顺 meek_133,无声无息 silent/quiet_126, 贴心 intimate_116, 优雅 elegant/graceful_113, 懒 lazy_112, 蜷 curl_109
天使 angel	291	6461	堕落 fall_1902,美丽 beautiful_1766,守护 guard_302, 飞翔 fly_301, 可爱 lovely_296,飞 fly_241,纯洁 pure_188, 坠落 fall_72,美好 beautiful_67,漂亮 beautiful_59
狼 wolf	493	6062	嚎叫 howl_792,凶狠 fierce_699,战斗 fight_450, 思考 think_310, 嚎 howl_262, 扑 rush/attack_143, 阴狠 baleful_142, 牢牢守住目标 hold the target_136,叫 howl_102,恶 fierce_99
阳光 sunshine	286	4987	温暖 warm_1389, 灿烂 bright/shining_986,包围 surround_562, 照耀 shine_296, 普照 shine_148, 洒 shine_136, 明媚 sunny/shining_127, 耀眼 radiant/glare_106,透明 clear_101,照亮 shine_63

Table 4. Top 10 Most Frequent Words in CogBank

Table 5 shows the most frequent properties in CogBank: *beautiful*, *bloom*, *fight*, *fly*, *convenient*, *warm*, and *painful*. Each property is associated with hundreds of words. But the frequency of the top 10 concept words occupies more than half the occurrences. This indicates that people is tend to use the same kinds of vehicles to form a simile expression.

Prop	# of Words	Freq	Top 10 Words
美丽 beautiful	816	17383	自由 free_3285,天使 angel_1766,花儿 flower_1202,花 flower_864,美玉 jade_843, 嫦娥 Chang E_795, 天神 god_342, 凤凰羽毛 phoenix feather_283,彩虹 rainbow_260,首都金边 Phnom Penh_242
绽放 bloom	152	16150	花儿 flower_7809,花 flower_6419,花朵 bloom_269,鲜花 flower_235,莲花 lotus_149, 玫瑰 rose_108, 昙花 epiphyllum_106, 蓝玫瑰 blue rose_85, 烟花 fireworks_76, 玫瑰花 rose_57
战斗 fight	217	13536	男人 man_8771,英雄 hero_547,艾薇儿 Avril_473, 狼 wolf_450, 战士 soldier_295, 熊 bear_229, 爷们 menfolk_145,保尔 Pual_118,斯巴达克 Spartacus_108,勇士 warrior_99
飞 fly	375	12409	鸡毛 chicken feather_2298, 子弹 bullet_1427, 蝴蝶 butterfly_890, 鸟 bird_769, 小鸟 birdie/dickey_657, 箭 arrow_522, 鸟儿 bird_453, 风筝 kite_380,叶子 leaf/foliage_372,雪片 snowflake_322
简单 simple	916	8133	涂指甲油 nail polish_757, 火焰 flame_328,呼吸 breathing_231,花儿 flower_220, 打开冰箱 open the fridge_200, 吃饭 eat_188, 拉屎 shit_138, 骑单车 cycling_131, 遛狗 walk the dog_118,孩子 child/kid_115
盛开 bloom	68	6970	花 flower_5375,花儿 flower_681,鲜花 flower_259, 蔷薇 rose_105, 花朵 bloom_96,烟花 fireworks_72,向日葵 sunflower_32, 桃花 peach blos-

			som_30, 樱花 sakura_26, 恶之花 flowers of evil_24
方便 convenient	625	5988	存款 deposit_388,电脑登录 login by computer_331, 控制电灯 control lamps_188,地铁 metro_143,取存款 withdraw_136,加油 refuel_131,取款 withdraw_129, 公交 bus_122, 家 home_118,公交车 bus_96
温暖 warm	289	5374	阳光 sunshine/sunlight_1389, 家 home_1207, 太阳 sun_535, 春天 spring_492,春风 spring breeze_132, 火炕 heated kang_109,爱情 love_98, 火 fire_77, 家庭 family_65, 拥抱 embrace/hug_61
痛 painful	479	5142	针扎 needle hit_1294, 抽筋 cramp_453,针刺 acupuncture_414,痛经 dysmenorrhea_314,刀割 cut with knife_284,散了架 fall apart_140,来月经 menstruate_102,死 die_98,火烧 burned_80,抽经 cramp_63
飞翔 fly	129	5014	鸟 bird_1290, 鸟儿 bird_883, 落叶 defoliation_505, 鹰 eagle_410, 小鸟 birdie/dickey_315,天使 angel_301,蝴蝶 butterfly_97, 飞鸟 bird_89, 雄鹰 eagle_70,风筝 kite_70

Table 5. Top 10 Most Frequent Properties in CogBank

4.2 Valuable Information from Uninformative Data

The manual data check drops many uninformative data which on the surface seem to possess no value, for example, "as stupid as ***you***", "as cheap as ***before***". The pronouns and time expressions have to be removed from CogBank. But through observing all the pronouns and time expressions through manual data check, we find something useful in Chinese sentences "X 像 (as) Y 一样 (same) A" (X is as A as Y) where Y is the reference object. As Indicated in Table 6, people prefer to use *我(I)* as the reference object rather than other pronouns.

Pronoun	# of Props	Freq
我 I	21962	54353
你 you	8422	20056
他 he	5678	12908
他们 they	3829	10315
她 she	2915	6576
我们 we	2583	5845
自己 self	1268	2537
别人 somebody else	1128	2291
其他人 others	1117	2437
你们 you pl.	1044	2124
它 it	519	1234
它们 they[-animate]	184	381

Table 6. Most Frequent Pronouns in Raw Data

People also prefer to reference recurring and concurrent time frames over past or future ones. As shown in Table 7, *usual (往常,平时)* occurs more than *past* and *before*, while *future(未来)* occurs with even lower frequencies.

Rank	Time	# of Props	Freq
1	往常 usual	18077	42895
2	现在 now	2320	5837
3	以往 before	2264	4563
4	从前 before	2263	4881
5	平时 usual	1705	3776
6	上次 last time	1584	3837
7	过去 past	1434	3431
8	今天 today	1124	2175
9	往年 years before	973	2179
10	往日 days before	775	1767
32	未来 future	19	557
37	明天 tomorrow	13	364

Table 7. Most Frequent Time Words in Raw Data

The usage patterns showing much higher frequencies for the pronoun *我(I)* and time expression *往常, 平时(usual)* suggest that people prefer to use their experienced everyday life knowledge to make simile or contrast sentences. This finding supports the Embodied Cognition Philosophy (Lakoff & Johnson 1980; Lakoff 2008), which hypothesizes that much of our conceptual structure is based on knowledge formed through physical and emotional experience.

Work on this kind of knowledge is still in its preliminary stage, and presents the potential to advance smarter automatic metaphor generation and QA systems.

5 Online Search and Visualization

The web version[1] of Chinese CogBank provides basic and visualized searches. Users can search for the properties of a particular concept or the concepts associated with a specific property. We also developed a search service for English users. An English word like *snow* will be translated into Chinese first with HowNet, and then the system will show its properties with English translations.

The above search services are provided in ordinary table form. We also use the Visualization

[1] http://cognitivebase.com/

Toolkit D3 [2] (Bostock, 2011) to draw dynamic graphs for the search results. The functions are listed as follows.

(1) Generate the graph of properties for a given word. Or generate the graph of words for a given property.
(2) Generate a graph comparing properties for given words. Or generate a graph comparing words for given properties.
(3) Generate the extended graph of properties for a given word. The graph is extended by the words for the properties. Or generate the extended graph of words for a given property. The graph is extended by the properties for the words.
(4) Generate the graph of properties for a given word with sememes in HowNet.
(5) Generate the graph of properties for a given English word with translation by HowNet and extended by the sememes in HowNet.

Appendixes A-E illustrate the visualization graphs. Due to the copyright of HowNet, functions (4) and (5) have not been made available online. Many more visualization functions are currently under development. We hope these online services will help linguistic researchers and second language learners with their studies.

6 Discussion

Veale (2014) argues that such knowledge is useful for metaphor, irony, humor processing and sentiment extraction. The cognitive database with a large store of properties will be useful for both linguistics and NLP. Nevertheless, we still face many challenges in developing a metaphor processing system. We now discuss some of the problems in using such a resource in NLP.

(1) Cognitive properties cannot be used directly in simile and irony generation. It seems straight forwards but there are many complicated aspects of simile sentence generation. For example, if we want to generate a simile sentence to express that someone is very tall, we could simply query CogBank for the words having tall properties. Then we find words like *mountain, hill, tree, giraffe,* etc. We may say "Tom is as tall as a giraffe". But it's odd to say "Tom is as tall as a mountain" or "Tom is taller than a mountain" unless in fairy tales. However, when we want to express some building is very tall, we would choose mountain and hill but not giraffe. If we say "the building is as high as a giraffe", it is more likely to be an ironic statement. So it's obvious that the tenor in the sentence will influence or restrict the choice of vehicle. In simile generation, scientific world knowledge seems indispensable.

(2) Cognitive properties alone are not sufficient in metaphor understanding. If one says "Tom is a pig", we have to indicate whether it is a metaphor or not. If it is, the cognitive properties will supply the candidate ground of the metaphor. The problem is that there are so many properties that the ground may vary in different contexts. Sometimes it is "greedy", and sometimes it is "fat". Reconciling such ambiguity and contextual dependency requires a dynamic model for the context.

To sum up, there is still much work to be done before we are able to completely integrate cognitive word knowledge in language processing systems.

7 Conclusion and Future Work

In this paper, we introduced the construction of Chinese CogBank which contains 232,590 items of "word-property" pairs. Querying search engines with simile templates is a fast and efficient way to obtain a large number of candidate pairs. But to increase the quantity and quality of the database, manual check and adjudication are necessary. Using CogBank we identified interesting preferences people exhibit during production of similes in natural language. We also established multiple online search and visualization services for public use.

In the future, we will make further investigate of the CogBank's raw and labelled data. Second, we will compare the cognitive features across languages. Third, we will try to adapt CogBank for deployment in Chinese metaphor processing systems.

Acknowledgments

We are grateful for the extensive and constructive comments provided by the blind peer reviewers. We are especially thankful to Patricia Lichtenstein for the proofread revision. This work was supported in part by National Social Science Fund of China under contract 10CYY021, 11CYY030 and

[2] http://d3js.org/

Qing Lan Project Fund of Nanjing Normal University.

References

Dong, Z. D. & Dong, Q. 2006. *HowNet and the Computation of Meaning*. Singapore, World Scientific Press.

Fauconnier G., Turner. 2003. *The Way We Think: Conceptual Blending and The Mind's Hidden Complexities*. Basic Books.

Jia Y. X. and Yu S. W. 2009. Instance-based Metaphor Comprehension and Generation. *Computer Science*, 36(3):138-41.

Kintsch, W. 2000. Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review*, Vol. 7: 257-66.

Lakoff, G. & Johnson. M. 1980. *Metaphors We Live by*. Chicago: The University of Chicago Press.

Lakoff, G. 2008. *Philosophy in the Flesh*. Basic Books.

Bostock M., Ogievetsky V., Heer J. 2011. D3: Data-Driven Documents. *IEEE Trans. Visualization & Comp. Graphics*, Vol. 17(12):2301-2309.

Miller, G. A., R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller. 1990. WordNet: An online lexical database. *Int. J. Lexicograph*. 3, 4:235-244.

Roncero, C., Kennedy, J. M., and Smyth, R. 2006. Similes on the internet have explanations. *Psychonomic Bulletin and Review*, 13(1).

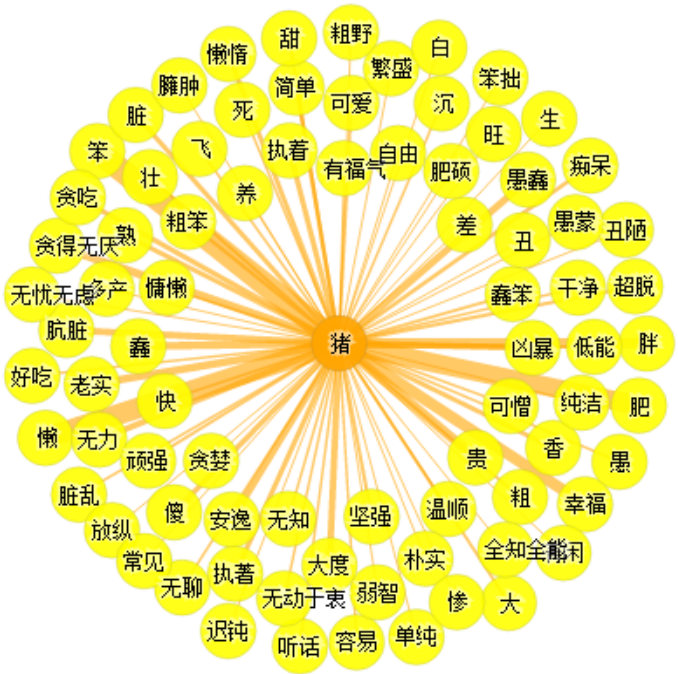
Veale, T. & Hao, Y. F. 2007. Learning to Understand Figurative Language: From Similes to Metaphors to Irony. *Proceedings of CogSci 2007*, Nashville, USA.

Veale, T., Hao, Y. and Li, G. 2008. Multilingual Harvesting of Cross-Cultural Stereotypes. *Proceedings of ACL 2008*, Ohio, USA.

Veale, T. 2014. A Service-Oriented Architecture for Metaphor Processing. In *Proceedings of the Second Workshop on Metaphor in NLP*, ACL 2014.

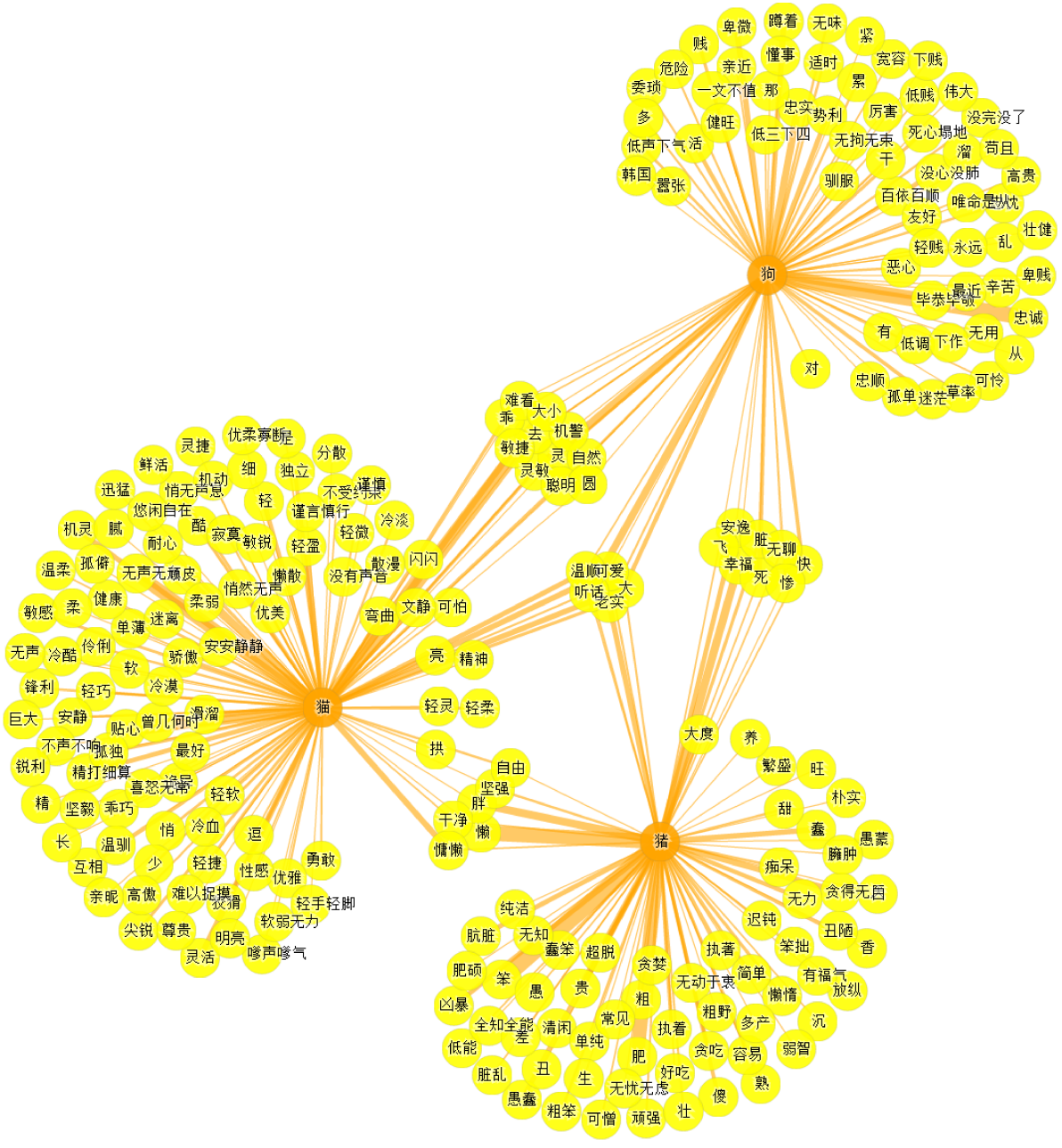
Appendix

A. The graph of properties of 猪(pig).



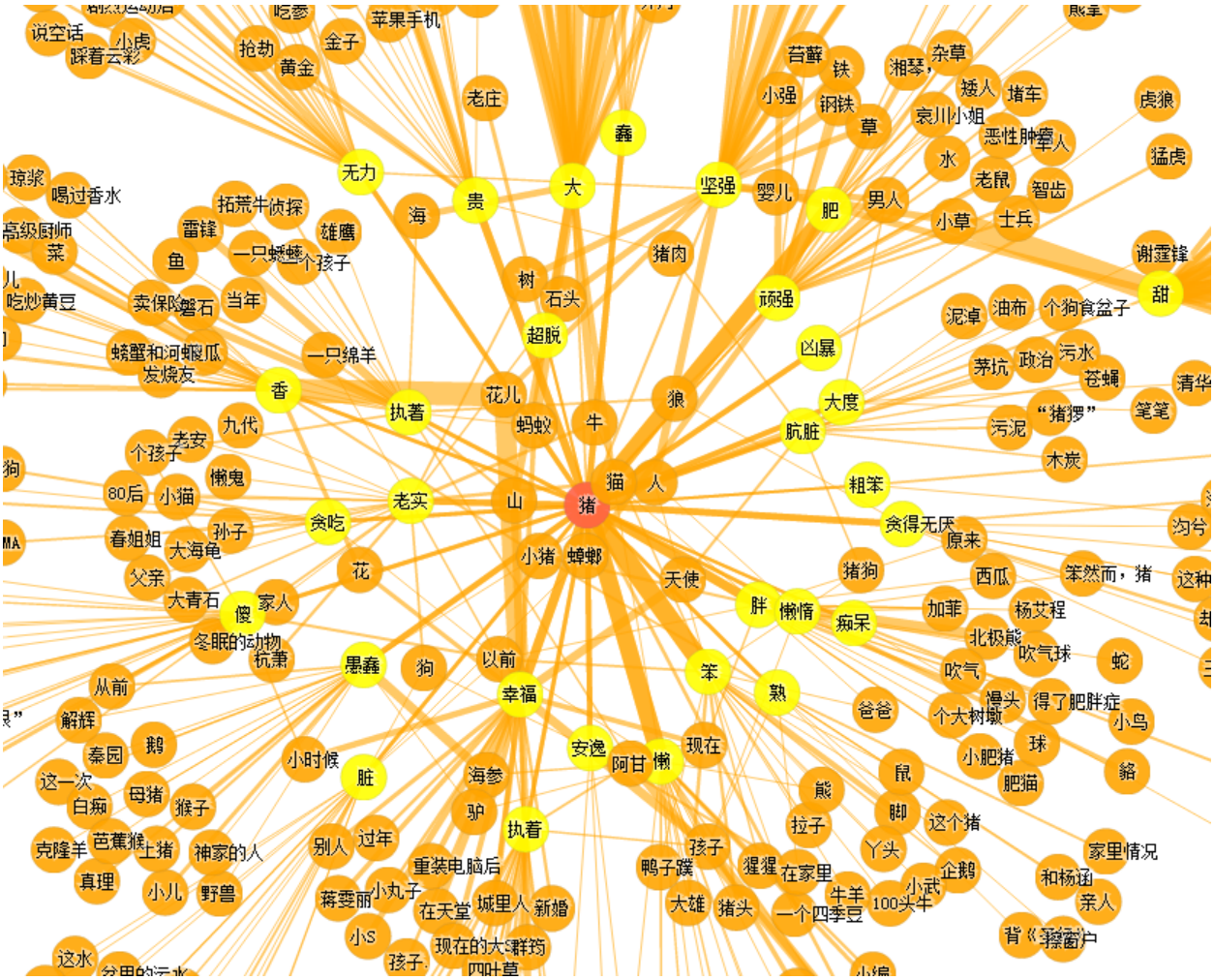
The word 猪(pig) is in the center surrounded by its properties.
The width of the line indicates the frenqucies of the "word-property" pairs.

B. The comparison graph of 猪(pig), 猫(cat) and 狗(dog).



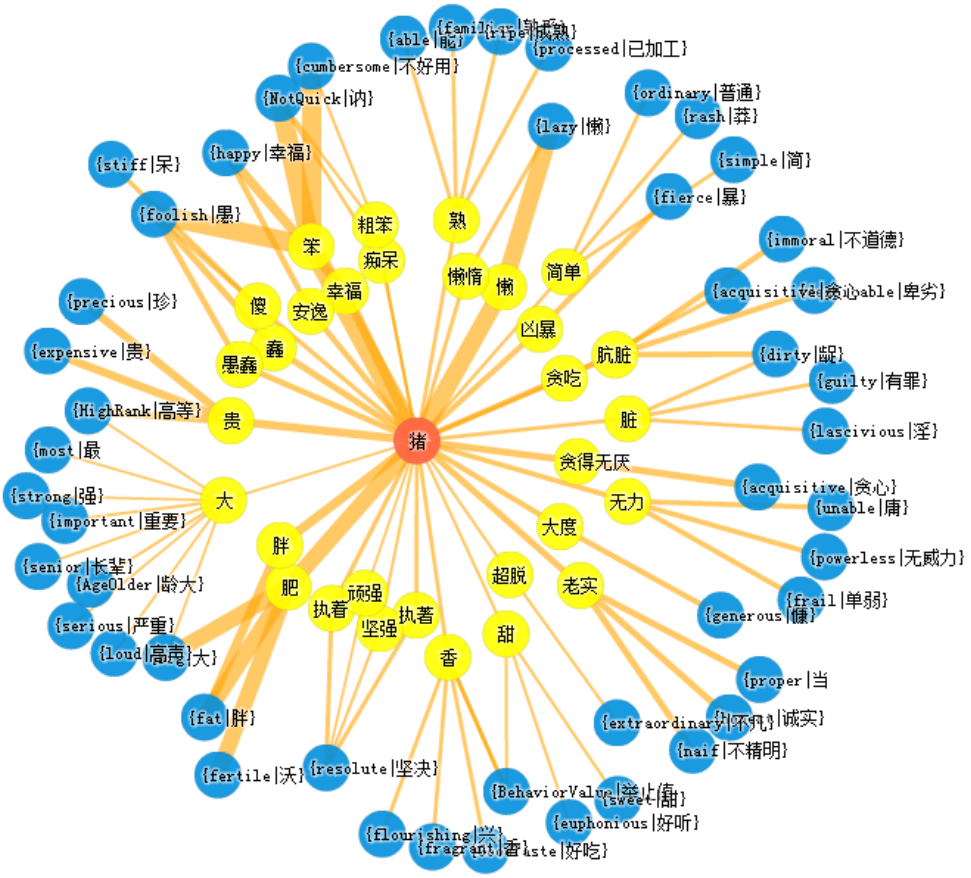
The 猪(pig), 猫(cat) and 狗(dog) share some properties.

C. The visualized graph of 猪(pig) with extensions by the propeties of 猪(pig).



The 猪(pig) has many properties(yellow nodes), and these yellow nodes have their words(orange nodes). So the extendted orange nodes share more properites with 猪(pig) appear closer to the center.

D. The visualized graph of 猪(pig) with bilingual sememe labels from HowNet.



The word 猪(pig) is in the center surrounded by its properties. Each property is linked to a bilingual sememe in HowNet (blue nodes).

E. The comparison graph of "sheep" with translation "羊" extended by HowNet's sememes.