Lexical Semantics-Syntactic Model for Defining and Subcategorizing Attribute Noun Class

Xiaopeng Bai, Hui Wang

Department of Chinese Studies, National University of Singapore
The Shaw Foundation Building, Block AS7, Level 3, 5 Arts Link, Singapore 117570, Singapore
{baixiaopeng, chswh}@nus.edu.sg

Abstract. Either word sense classification or word sense clustering can not avoid this problem: what features people can get from word meaning that can be used as the tools employed for producing word sense ontology. In this paper, we present a lexical semantic-syntactic model to define and subcategorize a noun class—attribute noun class, by acquiring syntactic behavior and semantic restrictions from corpus. First we introduce the idea of attribute noun class. Then we describe the model, which is a specific tool providing a practical method to decide whether a word should be located in a specific word sense class or not. Then we apply the model to 15 nouns to show how the model works, and create a sample taxonomy for the 15 nouns.

Keywords: attribute nouns, word sense classification, syntagmatic features

1. Attribute Noun and Attribute Noun Class

Concept (refers to either entity or abstract thing) has several characteristics viewed from different aspects. All things have properties: physical, social, etc. see the expressions below:

a1 *The color of the hat is red.*

a2 *这人脾气暴躁。(The guy's temper is irritable.)*

 "color" in a1 is a physical property of hat. "*脾气*(temper)" in a2 is a psychological property of people.

  Attribute noun class (as ANC below) refers to such characteristics/properties of things (like "*color*" of hat and "*脾气*(temper)*" of "*人*(man)" above). The noun is called attribute noun if it can be classified into ANC. The "attribute" presents the incomplete information—it is supposed to combine with other words to make an integrated expression, in other words, the attribute noun "looks forward to" collocating with words which can specify its meaning. In a1, "*red*" specify the meaning of attribute noun "*color*"; in a2, "*暴躁*(irritable)" specify the meaning of attribute noun "*脾气*(temper)"

2. The syntactic and semantic features of attribute noun

Syntagmatic mode of word has two sub-modes: syntactic behavior and semantic selectional restrictions. While the lexical meaning of word impacts the syntagmatic mode.

The meaning of a word might impact the syntactic behavior of the word. For example, attributive adjectives (also called non-predicative adjectives or as qu bie ci known in Chinese) in Chinese like "雪白(snowy white)" can not be syntactically used with adverb, so "非常(very)/adv 雪白(snowy white)/adj" is not a legal phrase in Chinese. Because there is a adverbial morpheme "雪(snowy)" in word "雪白(snowy white)" which means "snowy", the word merges the meaning of adverb into its lexical meaning. Attribute noun can not be attributive in attributive-head relation and rarely be isolate subject in subject-predicate relation which we will show in the following section.

The semantic selectional restrictions are semantic frame in which words should appear. The semantic frame is the grammatical relation with semantic constrains. For example, there are two grammatical slots in VO relation: verbal and objective. The meaning of object must semantically match with the meaning of the verb and vice versa, in other word, either verb or object has semantic "preferences" to the collocate words; such "preferences" is "selectional restrictions" in lexical semantics. The selectional restrictions also exist in semantic role frame, but we do not discuss the issue in this paper.

2.1 Syntactic behavior of ANC

ANC is one of the classes making the noun ontology. There have to be some formal features to distinguish ANC with other noun classes—we have to tell how ANC different with other noun classes. The syntactic behavior, as we discussed before, is the grammatical relations in which attribute noun could appear.

Generally, noun is grammatically united other words with following syntactic relations:

a.   Subject-predicate (SP) relation: 今天 星期一。*(Today is Monday.)* Noun can be either subject or predicate in Chinese.

b.   Verb-object (VO) relation: 今天 下午 打 篮球。*(Play basketball this afternoon.)* Noun can be object.

c.   Attributive-head (AH) relation: 皇帝 的 新装。*(The king's new dress.)* Noun can be either attributive or head.

Chinese is SVO language which is word-order-dependent, so we might acquire the syntactic behavior by studying the grammatical distribution of attribute noun in these grammatical relations.

The meaning of attribute noun presents incomplete information—it is supposed to combine with other words to make an integrated expression, in other words, the attribute noun "looks forward to" collocating with other words which can specify its meaning. According to this, attribute noun can not be isolate subject in SP. In sentence "今天 星期一(today is Monday)", the noun "今天" is an isolate subject in the sentence, it can be subject by itself without combining with other word. Because subject should be definite in Chinese and present integrated information, so attribute noun can not be isolate subject in a sentence. For example,

a3 *颜色 是 绿色 的* (color is green)

the meaning of a3 presents incomplete information. It is supposed to tell the color of what thing or who so that people can understand the expression. Thus the attribute noun "*颜色*(color)" can not be isolate subject in sentence like noun "今天(today)".

With the same reason, in AH relation, the attribute noun can not be the isolate attributive modifying head. Because the meaning of attribute noun needs to be specified by other words rather than specifying, thus it can be head modified by the attributive in AH. Thus, if we put a word before a3:

a4 *衣服 的 颜色 是 绿色 的* (color of the clothes is green)

the meaning of a4 is complete, we have a AH, in which attribute noun is head, to make a4 understandable.

Therefore, we have the syntactic behavior of ANC by studying its lexical meaning.

F1. The attribute noun can not be the isolate subject in SP.

F2. The attribute noun can not be the attributive in AH.

We will use F1 and F2 to determine whether a noun is an attribute noun or not in the section below.

Worth to mention that: 1) the rules of F1 and F2 is not black or white rule, the syntactic features occur with very big probability; 2) when we test whether a noun is an attribute one, the F1 and F2 should be applied in sentence level, i.e. if a noun is an isolate subject or attributive, the SP or AH must be semantically understandable as a sentence.

2.2 Semantic features of ANC

The ANC is a big word sense class and its meaning is general concept, thus we need to divide ANC into small sense classes in terms of more specific meaning. We use the semantic features to subcategorize ANC which are the semantic constrains between attribute noun and the collocate words in SP, VO and AH.

The selectional restrictions can be learned by studying the meaning of attribute noun and it's collocate words in AH, SP and VO.

2.2.1 Attribute Noun Used as Subject in SP

Predicate tells something about the subject. In SP, predicate is the value of attribute noun. Attribute nouns with different meanings calls for predicates that with different concept. Thus, we can assign an attribute noun to more specific ANC class in terms of the meanings of predicate collocated with.

2.2.2 Attribute noun as object in VO

Attribute noun is object in VO, which is one of the arguments of verb. Object semantically matches with verb as other types of arguments. For example, for the attribute noun "*外貌*", it can collocate with verb with the meaning of "watch" like "*看*" in VO (*看/v 他 的 外貌/o*, watch his appearance); while for the attribute noun "*性格*", the verb should be with the meaning of "mental action" like "*了解*" (*了解/v 他 性格/o*, know his temperament). Different attribute nouns "tend to" collocate with different verbs.

2.2.2 Attribute noun as head in AH

The function of attributive in AH is specifying the head. There are two meanings of attribute: 1) the value of attribute (*昂贵/a 的 价格/h*, high price) which semantically resembles the predicate in 2.2.1; 2) the host which semantically possess the attribute (*书/a 的 价格/h*, price of the book). Thus the meanings of host of attribute and attribute, as well as attribute value and attribute should match each other.

The attribute nouns can be assigned to more specific ANC classes in terms of the meanings of words with which the attribute nouns collocate in SP, VO and AH. Thus the selectional restrictions in three grammatical relations can be used as subcategorization criterion for ANC.

3. The lexical semantic-syntactic model for defining and subcategorizing attribute noun

Now we deduce a lexical semantic-syntactic model (LSS model) from above analysis for defining and subcategorizing attribute noun. We use pseudo-code of C to describe the model below:

```
Define AN
{
  syn_behvior;
  sr_SP;
  sr_VO;
  sr_AH;
}
AttrDef(AN noun)
{
  if(Be_Attr(noun, AH)==false||Be_IsoSub(noun, SP)==false)
    Subcat_AN(noun);
      else
        printf("the noun is not an attribute noun");
}
Subcat_AN(noun)
{
  while(sentence[i])
    features=get_syn_features(noun, sentence[i]);
  noun=SR(noun, features, AH)
    store the result in AN;
  noun=SR(noun, features, VO);
    store the result in AN;
  noun=SR(noun, features, SP);
    store the result in AN;
}
end.
```

The attribute noun is presented in a structure AN, in which the syntactic behavior and selectional restrictions are stored. The function AttrDef decides whether a word is an attribute noun or not, using the syntactic features F1 and F2. Then the word is input in Subcat_AN which assigns the word to a more specific sense class. In Subcat_AN, a noun is semantically analyzed in SR function with 3 syntactic relations: AH, VO and SP. The result of syntactic and semantic analysis will be presented in structure for each input.

4. Experiment

In this section, we make an experiment to evaluate the applicability of LSS model.

4.1 Methods

First we choose 15 nouns from Singapore Chinese Textbook Corpus. Then we abstract sentences from the first page result that CCL online corpus generated for each noun and acquire the data of syntactic features. The syntactic features are used to determine whether a noun is attribute noun or not.

Second, we choose 15 attribute nouns from the same corpus and generate sentences for each attribute noun from CCL online corpus, for abstracting semantic features. The semantic features will be put in structure AN mentioned in section 3 to show how 15 nouns differ in

semantic selectional restrictions. Such semantic features help us to divide 15 attribute nouns into several subcategories.

The features, either of syntactic or semantic, are manually abstracted from sample sentences. We do not employ machine learning software (like Stanford Parser, SRL software) to do the job, because the purpose of this paper is to create a model for attribute noun and evaluate rather than evaluate machine learning algorithm. We want the data to verify the model. In additional, the data set is quite small (15 words).

4.2 Syntactic behavior to determine whether a noun is attribute noun

The 15 nouns are: 老百姓(people), 华裔(foreign citizen of Chinese origin), 茶客 (customer), 矿长(mine manager), 珍珠(pearl), 甘泉(spring), 气浪(airsurge), 声音(sound), 乾坤(the universe), 前臂(forearm), 长度(length), 指标(indicator), 外貌(appearance), 性能 (performance), 直径(diameter). The last five nouns are already tagged as attribute noun. We use 4 indicators to show the grammatical distribution: isolate subject in SP, head in AH, Attributive in AH, object in VO. A word is called isolate subject when it is the subject of a sentence all alone, rather than in a noun phrase.

Table 1

Word	Isolate sub in SP	Head in AH	Attributive in AH	Object in VO
老百姓/45	13	9	10	13
华裔/50	0	0	50	0
茶客/50	22	18	4	6
矿长/50	2	23	23	2
珍珠/40	4	14	11	11
甘泉/43	6	18	2	17
气浪/	4	37	1	4
声音/49	5	33	1	10
乾坤/42	9	1	2	32
前臂/51	14	14	13	10
长度/39	0	34	3	2
指标/47	0	45	0	2
外貌/40	0	34	2	4
性能/42	0	41	0	1
直径/32	0	32	0	0

Table 1 shows the grammatical distribution of the 15 nouns. Five attribute nouns show quiet distinguish grammatical distribution from other nouns: they can not be isolate subject in SP and rarely be attributive in AH. This distinguishable grammatical distribution verifies that F1 and F2 rules are valid.

4.3 The 15 attribute nouns for abstracting semantic features

The 15 nouns are: 力气(strength), 天分(talent), 语调(tone), 生命力(vitality), 本能 (instinct), 歌喉(voice), 家境(family circumstances), 职务(position), 性格(character), 记忆 力(memory), 意图(intention), 嘴脸(look), 装束(garb), 直径(diameter), 风度(grace).

The nouns are randomly chosen from the corpus.

4.4 Selectional restrictions in AH

Table 2 shows the host word (noun) and attributes nouns (head) in AH.

<div align="center">Table 2</div>

Attribute nouns	Host words
力气 (strength, physical power)	李逵 (name of a man), 男人 (man), 蓝鲸 (blue whale), 蚂蚁 (ant), 小象 (baby elephant)
天分 (talent)	齐达内 (human's name), 学生 (student), 农民 (farmer), 女同志 (female colleague), 芬兰人 (Finland people)
语调 (tone)	朱镕基 (human's name), 外长 (minister of foreign affairs), 总理 (prime ministry), 他 (he), 老师 (teacher)
生命力 (vitality)	微生物 (microorganism), 植物 (plant), 虾 (shrimp), 蝈蝈 (katydid), 紫荆花 (Chinese redbud)
本能 (instinct)	动物 (animal), 鸟 (bird), 母豹 (female leopard), 人 (human)
歌喉 (voice)	俄耳甫斯 (human's name), 她 (she), 歌唱家 (singer)
家境 (family circumstances)	拉普拉斯 (human's name), 学生 (student), 贾玉峰 (human's name), 卢梭 (human's name)
职务 (work position)	他 (he), 文武百官 (government officers), 人物 (people), 中校 (lieutenant)
性格 (character)	老师 (teacher), 妈妈 (mother), 自己 (myself)
记忆力 (memory)	动物 (animal), 人 (human), 患者 (patient), 老年人 (old people), 蜜蜂 (bee)
意图 (intention)	导演 (director), 我 (I), 领导 (boss), 教练 (coach)
嘴脸 (look)	爪牙 (lackey), 政客 (politician), 美国人 (American), 敌人 (enemy)
装束 (garb)	妇女 (women), 姑娘们 (ladies), 她 (she), 哨兵 (guard), 将领 (general)
直径 (diameter)	银河系 (galaxy), 温室 (greenhouse), 手榴弹 (grenade), 坑 (pit)
风度 (grace)	绅士 (gentlemen), 学者 (scholar), 崔文华 (human's name), 女士 (lady), 政治家 (politician)

The hosts of attribute nouns "天分(talent)", "语调(tone)", "歌喉(voice)", "家境(family circumstances)", "职务(position)", "性格(character)", "意图(intention)", "嘴脸(look)", "装束(garb)" and "风度(grace)" are all of meaning of "human". The structure presentation of these nouns is:

天分(talent)={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[]};

语调(tone) ={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[]};

歌喉(voice)={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[]};

家境(family circumstances)= {syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[]};

职务(position)={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[]};

性格(character)={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[]};

意图(intention)={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[]};

嘴脸(look)={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[]};

装束(garb)={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[]};

风度(grace)={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[]};

→These nouns can be classified as "attribute of human".

The nouns "力气(strength)", "生命力(vitality)", "本能(instinct)" have hosts of not only the meaning of "human" also of "animal". Both meanings can be presented as "living creatures, thus the structure presentation of these nouns is:

力气(strength)={syn_behavior[]; sr_AH[living creatures]; sr_VO[]; sr_SP[]};

生命力(vitality)={syn_behavior[]; sr_AH[living creatures]; sr_VO[]; sr_SP[]};

本能(instinct)={syn_behavior[]; sr_AH[living creatures]; sr_VO[]; sr_SP[]};

→These nouns can be classified as "attribute of living creatures".

While the nouns above can be subcategorized by the meaning of the host words, while the noun "直径(diameter)" can not, because there is no semantic agreement of its hosts. The hosts are of various meanings, like "space" (银河系/galaxy), "architecture" (温室/greenhouse), and "artificial production" (手榴弹/grenade).

4.5 Selectional restrictions in VO

The ANC can be further classified by the verb with which the attribute noun matches. The attribute nouns and verbs are showed in table 3.

Table 3

Attribute noun	Verb
力气 (strength, physical power)	有 (have), 没有 (do not have), 下,费 (spend)
天分 (talent)	有 (have), 没有 (do not have),
语调 (tone)	运用 (handle), 用 (use)
生命力 (vitality)	显示 (show), 具有 (have), 保持 (maintain)
本能 (instinct)	有 (have), 成为 (become)
歌喉 (voice)	展示 (show), 听 (listen to)
家境 (family circumstances)	NULL
职务 (work position)	担任 (occupy), 开除 (dismiss), 辞去 (retire from)
性格 (character)	突出 (highlight), 表现 (present)
记忆力 (memory)	丧失 (lost), 保持 (maintain)
意图 (intention)	领会 (understand), 贯彻 (carry out), 知 (know), 揣摩 (conjecture)
嘴脸 (look)	描写 (describe), 揭露 (explosure), 看清 (realize), 露出 (reveal)
装束 (garb)	穿着 (wear), 看 (look), 瞅着 (stare at), 看清 (see clearly),
直径 (diameter)	NULL
风度 (grace)	注意 (pay attention to), 表现 (show), 保持 (maintain)

The verbs governing "*力气*(strength)", "*天分*(talent)", "*语调*(tone)", "*生命力*(vitality)", "*性格* (character)", "*本能* (instinct)", *记忆力* (memory) and "*风度*(grace)" are not specialized, the verb governing "*力气*(strength)" can also collocate with "*性格*(character)". Thus, we can not produce subcategories for these attribute nouns by the meaning of verbs.

The nouns "*嘴脸*(look)", "*装束*(garb)" and have verbs with meanings that only can collocate with them: "*看*(look)", "*瞅着*(stare at)". We can also find out some other verbs with same meaning for them: "*观察*(observe)", "*看见*(see)". If anything can be saw by human's eyes, watched or observed, it must has outward appearance. The structure presentation of these nouns is:

嘴脸(look)={syn_behavior[]; sr_AH[human]; sr_VO[watch]; sr_SP[]};

装束(garb)={syn_behavior[]; sr_AH[human]; sr_VO[watch]; sr_SP[]};

→Thus we create a new ANC subcategory for the two nouns: "appearance attribute of human".

The verbs of noun "*意图*(intention)" are of meaning of "mental processing", so it can be presented as:

意图(intention)={syn_behavior[]; sr_AH[human]; sr_VO[mental processing]; sr_SP[]};

→Thus we generate a subcategory for it: "mental attribute of human".

The meaning of verbs of "*职务*(position)" is "social acting", so it can be presented as :

职务(position)={syn_behavior[]; sr_AH[human]; sr_VO[social acting]; sr_SP[]};

→We create a new subcategory for it—"social attribute of human".

4.6 Selectional restrictions in SP

The nouns are needed to be analyzed in SP, for the AH and VO are not work on some nouns. Table 4 below is the attribute noun and predicates.

Table 4

Attribute noun	Predicate
力气 (strength, physical power)	很大 (very strong), 很小 (quite weak)
天分 (talent)	颇高 (pretty high)
语调 (tone)	低沉 (low), 平和 (gentle), 柔和 (soft), 平缓 (gentle)
生命力 (vitality)	旺盛 (vigorous), 强 (strong)
本能 (instinct)	NULL
歌喉 (voice)	嘹亮 (loud and clear), 极富磁性 (very charming), 宛转 (agreeable)
家境 (family circumstances)	贫寒 (poor), 富裕 (rich)
职务 (work position)	很高 (pretty high), 被升高 (be prompted)
性格 (character)	外向 (optimistic), 好坏 (good or bad)
记忆力 (memory)	衰退 (decline), 很好 (very good)
意图 (intention)	是认真的 (is serious), 很明显 (very clear)
嘴脸 (look)	暴露无遗 (completely reveal), 乌黑 (black)
装束 (garb)	淳朴 (simple), 粗鄙 (vulgar), 高雅 (elegant)
直径 (diameter)	0.5—5 微米 (0.5-5 micrometers), 小于 20 千米 (less than 20 km)
风度 (grace)	翩翩 (elegant), 高雅 (elegant)

The meanings of predicates of attribute nouns "力气(strength)", "天分(talent)", "语调(tone)", "生命力(vitality)", "本能(instinct)" and "记忆力 (memory)" are not distinguishable. The predicates like "很大 (very strong)", "很小(quite weak)", "旺盛(vigorous)", "强(strong)" have general meanings in semantic, which can collocate with many different attribute nouns.

The meaning of predicates of "歌喉(voice)" are "tone of sound" which can not collocate with other 14 nouns, thus "歌喉(voice)" can be presented as:

歌喉(voice)={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[tone of sound]};

→We get a new subcategory: "audible attribute of human".

The meaning of predicates of "家境(family circumstances)" mean "social status" which can only collocate with the noun. "家境(family circumstances)":

家境(family circumstances)={syn_behavior[];sr_AH[human]; sr_VO[]; sr_SP[social status]};

→The noun can be assigned with a new subcategory: "social attribute of human".

Meaning of predicates of "风度(grace)" mean "appearance status" which can only collocate with the noun:

风度(grace)={syn_behavior[]; sr_AH[human]; sr_VO[]; sr_SP[appearance status]};

→The predicate of "风度(grace)" and the verb of "嘴脸(look)" and "装束(garb)" have similar in meaning, thus "风度(grace)" can be assigned in "appearance attribute of human".
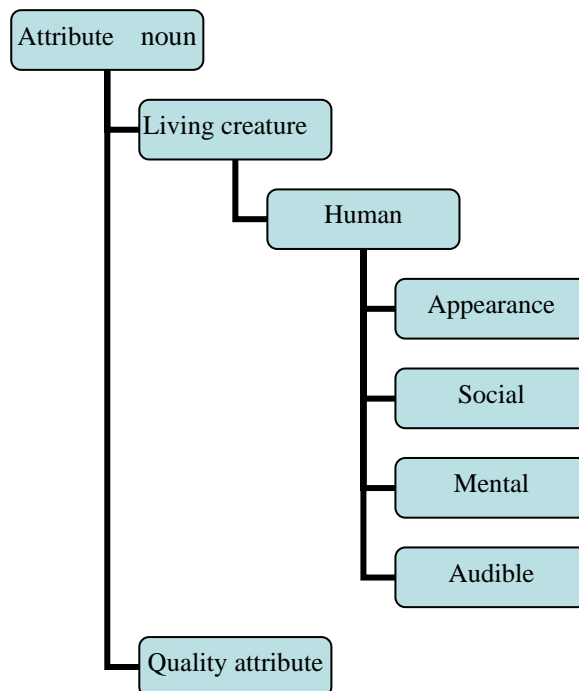
"直径(diameter)" takes quantifier phrase as its predicate in SP. It can be presented as:

直径(diameter)={syn_behavior[QP as predicate]; sr_AH[human]; sr_VO[]; sr_SP[quantity status]};

→So we get a subcategory: "quantity attribute".

4.7 Sample taxonomy

To summary, we create a sample taxonomy for these 15 attribute nouns:

5. Conclusion

In this paper, we present a practical method to define and subcategorize the attribute noun class, and produce sample word sense taxonomy for 15 Chinese attribute nouns. Syntactic and semantic features are employed to do the job, from which we derive a syntactic-lexical semantic model for defining and subcategorizing the word sense class. First learn the grammatical distributions of each word from corpus to determine a noun is whether attribute noun or not. Then we study the selectional restrictions between the noun and its collocate words in each grammatical relation and assign the noun to more specific sense class. Finally produce a small word sense taxonomy. We found that:

1. we get the syntactic behavior of word by studying the meaning of the word, thus sense classes (meanings with same concept) could be distinguished by syntactic behavior;

2. creating word sense class requires both syntactic and semantic features which we call "syntagmatic features" of word;

3. general meaning of collocate words is essential for acquiring the semantic features of word sense class, which means some coarse and solid concepts of noun, verb and adjective are needed. Such coarse and solid concepts are the precondition for studying the selectional restrictions between collocate words.

References:

Adrienne Lehrer, *Semantic fields and lexical structure*, North-Holland Press, 1974

Beth Levin, *English verb classes and alternations: a preliminary investigation*, University of Chicago Press, 1993

Thomas R. Gruber, Toward Principles for the Design of Ontologies Used for Knowledge Sharing, *Formal Ontology in Conceptual Analysis and Knowledge Representation*, Kluwer Academic Publishers, 1993

Scott McDonald, Michael Ramscar, Testing the Distributional Hypothesis: The Influence of Context on Judgements of Semantic Similarity, Proceedings of the 23rd Annual Conference of the Cognitive Science Society, 2001

Wang Hui, *A Syntagmatic Study on Noun Senses in Contemporary Chinese*, Peking University Press, 2004

ames Pustejovsky, Towards a Generative Lexical Resource: The Brandeis Semantic Ontology, *Proceedings of the Fifth Language Resource and Evaluation Conference*, 2006