Recognition of Comparative Sentences Based on Syntactic and Semantic Rules-System

Bai Linnan，Hu Renfen，Liu Zhiying

Beijing Normal University

linnanbai@126.com

layout

- **Background**
- **Introduction**
- **Syntactic and semantic features**
- **Method**
- **Experiment and analysis**
- **Conclusion**

Background

- In the field of Teaching Chinese as a Foreign Language, abundant real corpus of comparative sentences are required to guide the research and practice of "teaching" and "learning" of comparative sentences.

- Automatic recognition of comparative sentences is very significant.

Introduction

- Chinese comparative sentences have too many syntactic forms. Except for the most representative "BI (比) sentence", there are many different syntactic forms expressing comparative meanings.

Examples

- 他们的等级比我们高。（their level is high**er** than us.）
- "在衣食住行四事之中，"食"的程度高于其余一切. (Among basic necessities of life, the food is **more** important than others.)"
- "我游得不好，没有你游得好. (I'm not good at swimming, I don't swim as good **as** you)".
- 你跟我妈妈一样高。(You are **as** tall **as** my mother.)
- 这个饭馆儿的菜是不错，但是不如我们杭州的饭馆儿。(The restaurant food is good, but not **as** good **as** our restaurant in Hangzhou.)

- In our work, we choose six categories of comparative sentences.

- "BI(比) sentence,

- You/MeiYou/YouMeiYou(有/没有/有没有)+……+A,

- Zui/Hai/Geng(最/还/更)+A,

- A+YU(于),

- Xiang/He/Gen/Tong(像/和/跟/同)+……+YiYang (一样) (+A)

- "BuRu" (不如) type.

Syntactic and semantic features

- we defined that comparative sentences should contain or imply five elements: <u>comparative subject</u> and <u>comparative object</u>, <u>comparative mark</u>, <u>comparative aspect</u> and <u>comparative result</u>. The most important elements in comparative sentence are **comparative marks** and **comparative result words**.

Comparative marks

Table1. Comparative marks

HNC symbol	The marks	Example sentence
L0	BI, BUBI, YOU, MEIYOU (比，不比，有，没有)	
EU	GENG, HAI, ZUI (更，最)	
HV	YU(于), GUO (过)...	
		necessities of life, the food is more important than others.)
L1	XIANG, HE, GEN, TONG...(像，和，跟，同，与)	你跟我妈妈一样高。(You are as tall as my mother.)
EG	BURU...(不如)	这个饭馆儿的菜是不错，但是不如我们杭州的饭馆儿。(The restaurant food is good, but not as good as our restaurant in Hangzhou.)

The classification is based on HNC theory, these five marks are spread over different syntactic position.

comparative result

Table2 types of comparative result

types	illustration	examples
Adjective	Adjective predicate	你当然比我知道的多。(You certainly know more than me.)
Adjective + complement	The complement generally is HV or quantitative phrase	他们的水平比我们高多了。(Their level is much higher than us.)
VP	The verb is psychological verbs or modal verbs, phrases such like "一点儿/一些/得多" could be added after them.	我比你多了解一点儿情况。(I know more about the situation than you.)
(VP+)V+得+NP+A(OR: X+V+得+W+Y+A)	The verb is the common verb, adjective complement could be added after them.	她英语说得比普通话好。(She speaks English better than Manderin.)
Verbs meaning increase or decrease + the number of complement	The verb means increasing or decreasing, the following quantitative phrases or noun phrases are object.	每亩产量比去年增加了五十斤。(The yield per acre increased five pounds than last year.)
先/后/早/晚/多/少+V+ the number of complement	Patient object could be added.	我只不过比你早一点儿来到中国。(I'm just a little earlier than you come to China.)

Categories

Table3. Categories of comparative sentence

RM	A	A + complement	VP	(VP+)V+得 +NP+A(OR: X+V+得 +W+Y+A)	increase or decrease verbs	先/后/早/晚多/少+V+ the number of complement
BI(比)	√	√	√	√	√	√
BUBI(不比)	√	√	√	√	no	no
YOU(有)	√	no	√	√	no	No
MEIYOU(没有)	√	√	√	√	√	√
GENG(更/还/最)	√	√	√	√	no	no
YU(于/过)	√	no	no	no	no	no
XIANG(像/和/跟/同)	√	√	√	√	no	no
BURU(不如)	√	√	√	√	√	no

Method

- We build 4 models （L0 model，ABK model，EG model and QE&HV模型）to identify different types of comparative sentences.

- L0 model: 比、有

- ABK model: 跟……一样+adjective

- EG model：不如

- QE&HV model：更（好）、（高）于

L0 model

- **This is the most typical model of comparative sentences identification, and this model mainly process two types of comparative sentences, that are "BI" sentence and "You/MeiYou/YouMeiYou (有/没有/有没有)+……+A" structure. We used logical conception L0 to label "比/不比/有/没有/有没有", which means these conception could introduce main semantic chunk.**

- **Example sentence:中秋节有春节那么热闹吗？ (Is Mid Autumn Festival as lively as Spring Festival?)**

Phase1: exclusionary rules

- **Rule:**(0){CHN[有,没有,有没有]&LC_CC[v]}+(f){(1)CHN[这么,那么]+(2)LC_CC[u]}=> !LC_SELECT(0,LC_CC,v)&!LC_SELECT(0,LC_CC,jlu)&!LC_SELECT(0,LC_CC,hv)$

- In this sentence, comparative mark is the activating point, which is set as node 0 in our system. CHN means Chinese string. If there is an adjective at the location closest to node 0 when searching forward, and a Chinese string "这么" or "那么" adjoining to this adjective before. Attributes except L0 of "有" will be eliminated.

Phase2: generating the core predicate

- (b){(-2)CHN[有,没有,有没有]}+(-1) CHN[这么,那么]+ (0){LC_CC[u]&END%}=>LC_TREE(E,0,0)&PUT(fp,LC_E_SCORE, E_U)$

- An adjective is at the end of a sentence, we could search Chinese string "这么" or "那么" next to it ahead, and there is "有/没有/有没有"   before them. Then the adjective will be generated into an E tree, and the E was given the highest weight E_U.

Phase3: generating L0

- (0){CHN[有,没有,]&LC_CC[l02,l0]}+(f){(1)END %&LC_CHK[E]&LC_E_SCORE[E_U]}=>LC_TREE(L0,0,0)&PUT(fp,LEVEL,1)$

- In this sentence, "有" will be generated into a L0 tree whose level is 1, which means that it is the L0 of a sentence.

Pahse4: matching L0 and core predicate

- (b){(-1)CHN[有,没有,有没有]&LC_CHK[L0]&!LEVEL[2]}+(0){LC_CHK[E]&LC_E_SCORE[E_U]&END%}=>PUT(-1,LEVEL,1)+PUT(0,LC_E_SCORE,E_FORMAT)$

- After generating L0 and E, there is one essential step to finally identify the comparative. That is weighted adjective predicative as E_FORMAT by L0. Only in this way, this adjective could truly be core predicate.

ABK model

- structure :Xiang/He/Gen/Tong(像/和/跟/同)+……+YiYang/ChaBuDuo(一样/差不多) (+A)"
- Rule:(b){(-1)CHN[与,和,同,跟]}+(0)LC_CHK[L1H]&CHN[相比,相比较,一样]=>LC_TREE(L1,-1,-1)+LC_TREE(ABK,-1,0)$

EG model

- L0 model and ABK model mainly process comparative sentence with preposition comparative mark. There are some sentences using verb to express comparative relation, such as "不如/不及" and so on. This model is easier than the two models above. Firstly, these words have only one conception category. Secondly, the EG phase is the key step to identify comparative sentence, which is also the necessary step for every sentence.

QE&HV model

- The structure "A+于/过" and "还/更/最+A" are special types of comparative sentence. Because there is no need to set special phases to process these kind of comparative sentences, the procedure is part of EG phase. "过/于" and "还/更/最" could been seen as the affix of the front adjective. "过/于" is HV of the core predicate, which stands attribute component after verbs, while "还/更/最" are adverbs before verbs, and we use EU to label them.

Experiment

- **Data set :**

- Our test data come from the textbooks of Teaching Chinese as a Second Language.

- We have 12403 entries in our words library. The keywords are from the sentence we labeled, and they are also compiled into words library based on our labeling specification as shown before.

Table4. the result of identificatioon

Categories of comparative sentences	Precision	Recall
BI、BUBI+……+A	92.31%	82.07%
YOU、MEIYOU、YOUMEIYOU+……+A	92.98%	88.33%
XIANG、HE、GEN、TONG+……+YIYANG、CHABUDUO（+A）	98.11%	88.14%
ZUI、HAI、GENG+A	96.56%	89.58%
A+YU	96.88%	81.49%
"BURU"	96.72%	82.87%

Table5. the result of Comparative experiment

	Precision	Recall	F-score
SS+DR+SVM	85.4%	88.2%	86.8%
Keywords, Entity, SCR	96.55%	88.63%	92.43%
HNC-system	95.59%	85.41%	90.22%

Analysis

- From table 4 we can see that our system has achieved the high Precision and Recall in every categories. From table 5 we can see that our system could achieve the same or better result compared to statistical method.

- Two reasons can account for the bad result.

- (1) Word segmentation and the wordlist have an influence on the identification result.

- (2) The corpus wordlist are not abundant for analysis all semantic rules.

Conclusion and future work

- This paper proposed strategy based on HNC rules of identifying comparative sentences in the field of TCSL. We give a clear categories based on syntactic feature and the different models to identify every type of them. Our semantic system has high precision and better stability. Our system is not only helpful to TCSL, but also available to future research about comparative sentences. In our future work, we will expand our data sets and improve the precision of our system. On this base, we will do comparative relations extraction.

THANK YOU!