

# 日中机器翻译中汉语副词的数据处理

张颖

(日本岐阜大学, 岐阜市, 日本, 501-1193)

**摘要:** 日中机器翻译需要大量的基于语法和概念标准的结构转换数据,为了使其控制结构更清晰化以及更于理解,本文通过本研究室研究的基于模式字典的数据引擎 JAW,根据对两种不同语言副词用法的不同,分析了与汉语副词相对应的日语表达方式,通过引擎 JAW 在其内部做好相对应的规则,使得机器能够根据既定的条件,选择合适的意思,然后得出符合人性化的翻译结果.

**关键词:** 机器翻译; 数据转化; 引擎 JAW; 副词; 模式

## Chinese Adverb Processing in Japanese-to-Chinese Machine Translation

Zhang Ying

(Gifu University, Gifu City, Japan, 501-1193)

**Abstract:** Japanese-Chinese machine translation requires a large amount of structural transformations in both grammatical and conceptual level. In order to make its control structure clearer and more understandable, this paper proposes a model ---Engine JAW, which based on pattern dictionary. This paper analyzes Japanese expressions which correspond to Chinese adverbs. We classify the expressions into two types which are translated differently. An adverb processing method in J-to-C machine translation using this classification is proposed. By using the Engine JAW, we make the corresponding rules to the different languages in order to let the machine know how to select the right words to make the right translation results.

**Key words:** Machine translation; Data transformation; Engine JAW; Adverb; Pattern

### 1. Introduction

#### 1.1. On Machine Translation

The type of MT is that of the 'text to text' variety. MT can be divided into two types, Unassisted MT and Assisted MT. Unassisted MT takes pieces of text and translates them into output for immediate use with no human involvement. The result is unpolished text and gives only a gist of the source, hence the term 'gisting'. The ultimate aim of this type of MT is sometimes known as Fully Automatic High Quality Translation (FAHQQT), perfect translation created solely by

---

作者简介: 张颖(1965. 10. 21), 男, 籍贯: 沈阳, 职称: 翻译(南昌科技大学), 在读学历: 博士(日本岐阜大学)

a computer.

Assisted MT uses a human translator to clean up after, and sometimes before, translation in order to get better quality results. Usually the process is improved by limiting the vocabulary through use of a dictionary and the types of sentences/grammar allowed. The use of a 'controlled language' has been fairly successful. Some systems have also been set up to learn from corrections. Assisted MT can be divided into Human Aided Machine Translation (HAMT), a machine that uses human help, and Machine Aided Human Translation (MAHT), a human that uses machine help. Computer Aided Translation (CAT) is a more recent form of MAHT. Another area of MT that is worth mentioning here is Natural Language Processing (NLP). NLP parses sentences and determines their underlying meaning in order for databases to answer SQL queries entered in the form of a question. For further information on the structure of MT systems see the recent special report on the future of translation featured in Wired magazine.

### 1.2. About the Engine JAW:

The Engine JAW, one kind of PC (Pattern Conversion) type Machine Translation engine, is using original language pattern to make the pattern dictionary. Making use of this dictionary, it expresses the input command with the IT (Input Tree) pattern. The feature of JAW is divided into three stages: the translation for proposition content; the translation for functional word after Japanese YOUGEN (Declinable); and the translation for functional word after Japanese TAIGEN (Nominals). (TAIGEN refers to words that don't inflect in any way (mainly nouns), and YOUGEN to words that do inflect.) JAW/Chinese is a machine translation system from Japanese into Chinese. JAW is the translation engine from Japanese to other languages; it means Japanese to Asian and World languages.

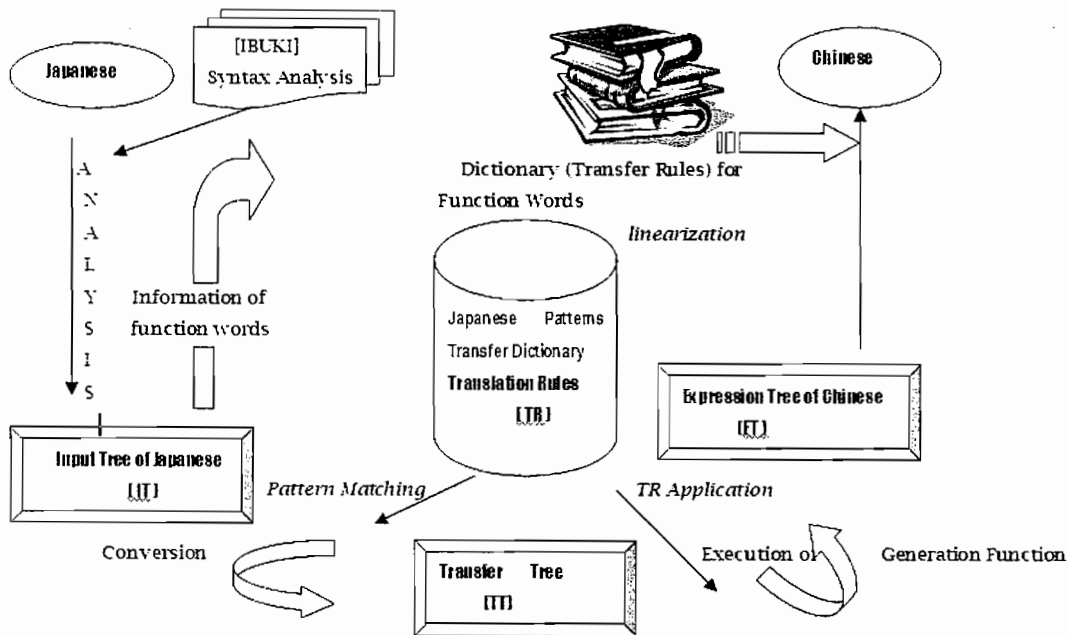


Figure : Outline of JAW/Chinese

After analysis of a Japanese sentence by IBUKI ( a system developed in our laboratory for segmentation of Japanese into bunsetsus), JAW put them into a tree (input tree: IT), then JAW search Japanese patterns in the transfer dictionary for the IT and make a tree of transfer rules (transfer tree: TT). The system is implemented on C++ and the transfer rule is in fact a C++ program stored as dll. The execution of the transfer rules in the tree produces a network of C++ objects for Chinese (expression tree: ET). A linearization function is defined for each object as a class method of C++. The execution of the linearization function on the ET, puts the members of ET in a line to make a Chinese output sentence.

## 2. The problems in Adverb processing in Japanese to Chinese machine translation

Adverbs have various complex grammatical functions in sentences, In natural language processing, the study of adverbs has not developed very far to date, compared with nouns and verbs, because it was thought that adverbs do not construct the main parts of sentence meaning and have various complex grammatical functions in sentences. However, adverbs occur frequently and make important contributions to sentence meaning. Thus, the accurate processing

of Chinese adverbs is required for high quality machine translation. From a linguistic point of view, linguists have examined adverb grammatical functions and meanings in detail [Quirk et al., 1985]. Conjuncts and disjuncts, usually called sentence adverbs. Other studies by linguists include those which handle the meanings of specific adverbs, such as "even", "still" and "already". There are also studies of adverb position in Chinese in general and positions of specific adverbs. In Chinese the adverbs can be used only in front of the verbs or adjectives. But it is difficult to apply the results of these studies to natural language processing directly because they are the knowledge for human so computers cannot understand them easily. From the natural language processing viewpoint, few studies have considered adverbs in natural language processing; it aimed to decrease ambiguity in adverb meanings and to select words during generation by applying information about adverb semantics and syntax from linguistic studies to an adverb lexicon. These are studies of the multiplicity of adverb meanings. A method for determining where adverbs should be placed in Chinese sentences in Japanese-to-Chinese machine translation has been proposed.

The main problems in adverb processing in Japanese to Chinese machine translation are the two as follows:

- 1) the multiplicity of adverb meanings
- 2) word ordering of Chinese adverbs in Chinese generation

The method is based on adverb grammatical functions (subjuncts, adjuncts, disjuncts and conjuncts) and meanings (process, space, time etc.), preferred positions in sentences (initial, medial, end, pre, post), and priorities between adverbs with the same preferred position. There are few studies of differences in expression between Japanese and Chinese for adverbial meaning. It showed that only about 55% of Japanese adverbs were translated into Chinese adverbs in translation from Japanese to Chinese by human translators. On the other hand, only about 17% of Chinese adverbs that appeared in the human translation were translated from Japanese adverbs in the original. This shows clearly the difference between Japanese and Chinese representations for adverbial meaning and the difficulty of adverb processing in machine translation. Adverb processing in Japanese-to-Chinese machine translation is very complicated. Thus, in this paper, we focus our attention on the problem of differences in expressions between Japanese and Chinese adverbial meaning from the viewpoint of Chinese adverbs. When translating between different language families, correspondence between language elements is not straightforward. The tendency is especially prominent in translation of adverbial expressions.

So first we examine in detail examples in which Japanese expressions correspond to Chinese adverbs. We classify the examples into two types, which are translated differently. An adverb processing method in Japanese-to-Chinese machine translation using the correspondence types and functional differences of adverb expressions is proposed. The content of Japanese dictionaries which are used to determine Japanese composition of words, Japanese-to-Chinese word transfer dictionaries which are used to transfer by word units, and Japanese-to-Chinese structure transfer dictionaries which are used to transfer by predicate unit, are also presented. Here, the pattern dictionary we used is provided by NTT Group (Nippon Telegraph and Telephone Corporation). At the same time we also use the Semantic Attribute Dictionary which shows the semantic relationship of words, and the Function Condition Dictionary which shows the condition for choosing the function words. These two dictionaries are from the Relational Data Base System provided by Microsoft Corporation.

### 3. Rules making in JAW/Chinese

To aim at this problem, we can make the different Japanese patterns to be match along with the Chinese translations as follows:

For example:

Japanese "mô もう"

①

Japanese Pattern (JP): mô(もう)+ ichido(一度) V suru(する)  
 Chinese Translation (CT): zai(再) V yici(一次)

example:

この映画は先週一度見たが、明日もう一度見ようと思う

kono eiga wa sensyu ichido mita ga, asu mô ichido miyô to omô.

CT: 这个电影我上周看了一次，明天想再看一次

zhege dianying wo shangzhou kan le yici, mingtian xiang zai kan yici.

②

Japanese Pattern: mô(もう)+ ichido(一度) V shita(した)  
 Chinese Translation: you(又) V le(了)yici(一次)

example:

この映画は先週一度見たが、昨日もう一度みた

kono eiga wa sensyu ichdo mita ga, kinô mô ichdo mita.

CT: 这个电影我上周看了一次, 昨天又看了一次  
zhege dianying wo shangzhou kan le yici, zuotian you kan le yici.

Japanese Pattern: mô kore ijyô (もうこれ以上)...nai(shinai)ない(しない)  
Chinese Translation: zai(再)+ ye(也)+ bu(不)/meiyou(没有)...

example: こういうチャンスは恐らくもうこれ以上ないだろう。

kô iu tyannsu wa osoraku mô kore ijyô nai darô.

CT: 这种机会也许再也不会有了。(也许再也没有了)。

zhezong jihui yexu zaiye bu hui you le. (yexu zai ye meiyou le).

Japanese Pattern: mô (もう)+ suryô meishi(数量名詞)+ V suru(する)  
Chinese Translation: zai(再) V suryou meishi(数量名詞)

example: もう一時間待って、彼が来なかつたら先に行く。

mô ichi jikan matte, kare ga konakataru sakini iku.

CT: 再等1小时, 他不来的话, 我就先走了。

zai deng yi xiaoshi, ta bu lai de hua, wo jiu xian zou le.

Japanese Pattern: mô (もう)N jikan/nenrei (時間/年齢) / V / A dj  
Chinese Translation: yijing/dou(已经 / 都)

example: もう8時だよ、起きなさい。

mô hachi ji da yo, okina sai.

CT: 已经/都8点了, 起来吧。

yijing/dou ba dian le, qilai ba.

Japanese Pattern: mô (もう) V deshô (でしょう)  
Chinese Translation: kuaiyao(快要)

example: もう来るでしょう。

mô kuru deshô.

CT: 快要来了吧。

kuaiyao lai le ba.

#### 4. Actualization

The basic CLASS (CDeclinableWord: CObject) for YOUGEN in JAW/Chinese is designed like this:

//動詞、形容詞の基本クラス(形容詞、動詞などのクラスにはすべてこのメンバが含まれます)

CAverb \*m\_verb 副詞

//「とても」美しい花。「とても」早く走る。「いつも」公園へ行く。

CNoun \*m\_nounModifier 用言に係るその他の修飾要素

//名詞、又は名詞を使うフレーズで用言を修飾する

//「年のわりに」若く見える。「公園で」遊ぶ。「明日」行く。

//どういふものが nounModifier になるかは後で説明します。

Adverb Class (CAdverb:CObject) is designed like this:

```
//このクラスの設計は他の言語に合わせていません
CString  m_evaluative  evaluative adv.  //「必ず」「らしい」「てみる」
CString  m_connective  connective adv.  //接続表現からの訳の副詞
CString  m_time        time adv.        //「いつか」
CString  m_freq        frequent adv.    //「大抵」「がちだ」
CString  m_scope       scope adv.       //「しかない」「みんな」
CString  m_degree      degree adv.      //「とても」「非常に」"掲械
CString  m_cooperative cooperative adv. //「一緒に」
CString  m_repeated    repeated adv.    //「再び」「また」
CString  m_descriptive descriptive adv. //「お互いに」「わざと」
CString  m_negative    negative adv.    //機能語から //艶、音、短
```

5. The real example for adverb translation by JAW/Chinese:

Japanese: 近年交通事故はますます増えている。

kinnen kōtsujiko wa masumasu fue teiru.

C-Translation: (SP\_teiru: -22):. 交通事故近几年<不断地><在 default>增加。

Figure 1 (Translation Pattern Editor---Japanese Pattern for Compound Word)

Figure 2 (Translation Pattern Editor--- Rule making for Compound Word)

規則登録済済 PatternID: 4300047 Title: 益々 キーワード: 益々 例:   
 RuleType: Addition 読み: ますます 英語:   
 属性: 属性 Comment:

日本語 English   
 Move   
 ±10   
 ±100   
 ±1000   
 Search   
 フィルター   
 ますます   
 Pid検索   
 検索条件クリア   
 新規登録

PatternID	PType	例	係り先文節番号	例	係り先文節番号	例	係り先文節番号	受身可能
4300047	1	益々	1	益々	2	益々	3	
4300047	2	ますます	1	ますます	2	ますます	3	

Figure 3 (Translation Pattern Editor--- Japanese Pattern for Adverb masumasu)

Microsoft Access - [Translation Pattern Editor]   
 ファイル 編集 表示 挿入 書式 レコード ツール ウィンドウ ヘルプ   
 規則登録済済 PatternID: 4300047 Title: 益々 キーワード: 益々 例:   
 RuleType: Addition 読み: ますます 英語:   
 属性: 属性 Comment:

Class	MemberName	MemberClass	Value
Proposition	m_adverb	Adverb	
Adverb	m_degree	String	不断地

Figure 4 (Translation Pattern Editor--- Rule making for Adverb masumasu)

規則登録済済 PatternID: 1114808 Title: 増える キーワード: 増える 例:   
 RuleType: Base 読み: ふえる 英語:   
 属性: 属性変化 属性変化 Comment:

日本語 English   
 Move   
 ±10   
 ±100   
 ±1000   
 Search   
 フィルター   
 増える   
 Pid検索   
 検索条件クリア   
 新規登録   
 追加・更新   
 削除

PatternID	PType	例	係り先文節番号	例	係り先文節番号	例	係り先文節番号	受身可能
1114808	1	増える	1	増える	2	増える	3	
1114808	2	ふえる	1	ふえる	2	ふえる	3	

Figure 5 (Translation Pattern Editor--- Japanese Pattern for Verb fueru)

規則登録済済 PatternID: 1114808 Title: 増える キーワード: 増える 例:   
 RuleType: Base 読み: ふえる 英語:   
 属性: 属性変化 属性変化 Comment:

Class	MemberName	MemberClass	Value	roleName	caseMarker
Proposition	m_centerW	String	増加		
Proposition	m_subject	Noun	2		
Proposition	m_nounModifier	Noun	1	time	

Figure 6 (Translation Pattern Editor--- Rule making for Verb fueru)

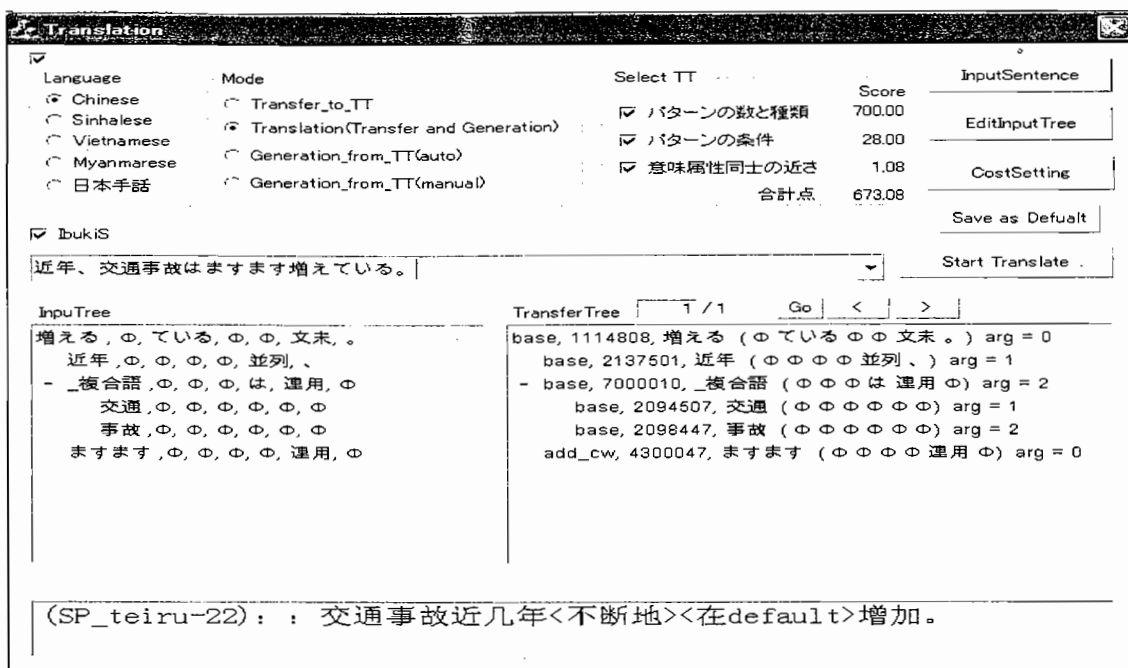


Figure 7(Translation Result including IT & TT by Analysis System---IbukiS)

Rule of Translation is designed like this:

	Rule Type	Class	Member
Adv	Addition_cw	CProposition	m_adverb
Noun	Base	CNoun	
Verb	Base	CProposition	
Adj	Addition_cw	CNoun	m_adjectival

Firstly, I have used our JAW/Editor to make Japanese Pattern for the Adverb masumasu like Figure 3 and to make a rule for it like Figure 4; and then to make Japanese Pattern and rule for the keyword---verb fueru like Figure 5 & 6, but the translation of this sentence didn't come out; I have tried to find the reason, I made the sentence short like this:

近年事故はますます増えている。  
 kinnen iko wa masumasu fue teiru.

And the translation of it comes out. From it, I found it is the problem of compound word---koutujiko. So I added a new pattern for the compound word like Figure1&2, finally I got the translation result like Figure 7.

Our experiment is performed on sentences with at least one Japanese adverb taken from the "1000 example Japanese sentences". It shows the translating meaning by a human translator. We manually examined whether Chinese adverbs in the translation would be generated correctly using the proposed method. By comparing with other software, we got the following result:

Result from the selling software:

1	日本人はよく働く国民である。	日本人是很好地工作的国民。
2	スイスは雪がたくさん降る。。	瑞士下很多雪的。。
3	彼はとうとう気が狂ってしまった。	他终于气被打乱了。
4	いま、おばさんの家に泊まっています	现在，阿姨的家停泊着。
5	近年、交通事故はますます増えている	近几年，交通事故越发增加着。

Result from our JAW:

日本人はよく働く国民である。	Start Translate
日本人是<经常地>劳动的国民。	
スイスは雪がなたくさん降る。	Start Translate
瑞士雪下<得><很多>。	
彼はとうとう気が狂ってしまった。	Start Translate
(SP_ta-13) : : 他<终于>发疯<了1>。	
いま、おばさんの家に泊まっています。	Start Translate
(SP_teiru-18) : : 现在在伯母的家住<着1>。	
近年、交通事故はますます増えている。	Start Translate
(SP_teiru-22) : : 交通事故近几年<不断地><在default>増加。	

Examined Objects

Japanese adverb entries: 86 words

Sentences: 122 sentences

Accuracy rate 92.3%

Adverbs generated in incorrect positions 21.05%

Absolutely incorrect position: 5.53%

Strange position: 15.52%

This experiment confirmed that the proposed word ordering method can handle large amount of adverbs correctly.

Our experiment had the Japanese to Chinese machine translation system JAW/Chinese translate Japanese sentences to test various Chinese adverb functions. The goal was to confirm that this adverb ordering method could handle various types of Chinese adverbs.

## 6. Word Ordering Method for Chinese Adverbs

Adverbs usually have many meanings, especially adverbs which are used frequently in our daily life. Normally the difference in meaning is indicated by the position in the sentence. The position of an adverb depends not only on the adverb's meaning but also on the relationship between the adverb and other sentence elements.

In Chinese, the basic order is :subject + adverb + verb/adjective + (verb complement) + object, here the Adverb is used in front of verbs or adjectives to show degree, extent, time or negation, etc., e.g. Degree: 很(very), 非常(very), 极其(extremely), 格外 (extraordinarily), Extent: 都(all), 仅仅(only) Time: 已经 (already), 曾经(ever), 刚刚 (just), 正在(at the moment), 立刻(immediately), 常常(often) Negation: 不(not), 没(no), 别(not) Positive: 必定(surely), 必(sure) Repetition or continuity: 又(again), 还 (again), 再(again) Mood: 却(however), 倒, 竟, 偏(even) In Chinese the adverbs can be used only in front of the verbs or adjectives while in English they may appear in front of the sentence.

The Chinese adverb ordering is very important in MT. The following are some examples including errors in the use of Chinese adverbs:

常常 (often)

Example : Wrong: 以前常常我去友谊商店买东西。 Right: 以前我常常去友谊商店买东西。

In Chinese, “常常” can not occur before the subject. One should put it between the subject “我” and the predicate verb “去”.

Example : Wrong: 现在我们跟中国同学常常说汉语。 Right: 现在我们常常跟中国同学说汉语。

The verbal predicate is preceded by the prepositional phrase “跟中国同学” which serves as an adverbial adjunct to introduce the object of the action “说”. The usual order in such a sentence is: the prepositional phrase goes directly before the verb with “常常” coming next.

Example : Wrong: 我不常常跟朋友一起吃饭。 Right: 我不常跟朋友一起吃饭。

The negative form of “常常” is “不常”. The second “常” is to be deleted.

Example : Wrong: 他常常早上在操场跑步了。 Right: 他常常早上在操场跑步。



“常常”indicates that an action occurs frequently in a certain period. The interjection“了”affirms a change of state. These two words do not match semantically. “了”must be deleted.

Example: Wrong: 他希望你以后常常也来玩儿。 Right: 他希望你以后也常常来玩儿。

“常常”and“也”are in the wrong order. When these words are used together,“也” meaning “also” normally goes before“常常”.

So different order can be led to different meaning or even wrong meaning in the machine translation from Japanese to Chinese. In our system we arrange the adverb according to the following order:

evaluative adv	connective adv	time adv	frequent adv	scope adv	negative
adv	cooperative adv	degree adv	repeated adv	descriptive adv	

## 7. Conclusion

A new adverb classification based on adverb grammatical functions, meanings and adverb preferred positions for Chinese adverb generation in machine translation was proposed. The effectiveness of the Chinese adverb ordering method in Japanese to Chinese machine translation based on the adverb classification and the priority of the same adverb preferred positions is shown. The priority was decided by examining sentences with adverbs to generate Chinese adverbs in Japanese to Chinese machine translation. If correct Chinese adverbs are selected after Japanese analysis and Japanese to Chinese transfer, about 97% of adverbs can be put correct positions in sentences by this adverb ordering method. The proposed method can also be used for adverbial prepositional phrases. In conclusion, Machine Translation can contain some errors and problems as we have read before, I have to say that in the last years Computers has improved and in future it will go on finding less problems .We have to say that language is very complex in order to have it translated: for instance, when you ask a program to translate you the word "manzanilla" it can may be answer you with "little apple" (obviously ,this does not always happens, it only happens when the translator is not good enough or when it does not have high quality).So, Computers is not able to register a whole language.

## Reference:

- Anyuan Luo "Simple Modern Chinese Grammar", Center Nationality University Press, (1996) P.145~153
- Conlon and Evans "Can computers handle adverbs?" In the 14<sup>th</sup> International Conference on Computational Linguistics, Proc.of COLING-92,NANTES, AUG. 23-28,1992, (1992) P1193~1195
- Hiromasa Imai "On Pattern Changing Engine for Machine Translation---JAW",  
Doctor Thesis of Ikeda Laboratory of Gifu University, (2004) P9,76~80,86~89
- Hirst, Graeme " Semantic Interpretation and the Resolution of Ambiguity---Studies in Natural Language Processing", Cambridge University Press, (1987)P184~186
- Matsuo Yoshihiro, Shirai Satoshi and Ikehara Satoru "Changing Syntactic Classes in Transfer-based Machine Translation", NLPRS 1995, Vol.1, (1995) P6~7
- Naoko Chino, Miharu Akimoto and Kazushi Sanada "Adverbs---Japanese For Foreigners", (1987) P1,2,5,75
- Shuiqing Shang "Japanese-Chinese Dictionary", Syougakukann(1992)
- Takuro Moriyama, Yoshio Nitta and Hiroshi Kudo, "Japanese Grammar---Modality", Iwanami Press, (2000)P164~231
- Yoshio Nitta "Phases of Adverb Expression", (2002) P33~41, P147,148, P265~266
- Yoshifumi Tobita,Hideko Asada " Modern Japanese Adverb Usage Dictionary", Tokyodoshoten Press (1994)
- Yuehua Liu, Wenyu Pan and Ye Gu "Chinese Grammar", Kurosio Press, (2000) P183~215