Classification of Modified Relationships in Japanese Patent Sentences

YOKOYAMA Shoichi
Yamagata University
4-3-16 Jonan, Yonezawa,
Yamagata, JAPAN 992-8510
yokoyama@yz.yamagata-u.ac.jp

KANEDA Yuya
Yamagata University
(now Yonezawa Credit Bank)

Abstract

It is well known that sentences in Japanese patents have long and complicated structures, especially necessary conditions and details. Here, patent sentences are analyzed and classified by pattern of modified relationships. Morphemes were first extracted using the famous morpheme analysis tool *Chasen*, and then the modified relations were extracted using the software *Cabocha*. Many modification mistakes were caused by long complicated structures, which required correction by humans. In the process of correction, the modification structure patterns were classified using about 200 sentences. This clarified the characteristics of Japanese patent sentences, and it is useful in machine translation of patent sentences.

Keywords: classification, modification, Japanese patent sentences, long sentences, complicated structure.

1 Introduction

It is well known that Japanese patent sentences have long and complicated structures with about 200 or more characters, amounting to 50 or 60 words (Figs. 1 and 2). Figure 1 shows an example of a Japanese patent form [1], and Fig.2 shows its translation by a person [2]; in the 'claim' section of the patent, translation into English was done using a machine translation system. As cited in References, all of the Japanese patent forms are published in both Japanese and English on the web. The sentences have long and complicated structure, especially in section 57, the 'Problem to be Solved', and in the 'Solution' section of the Abstract. Figures 1 and 2 also demonstrate that the construction of the modification is not clear.

This paper classifies the long and complicated structures in those patent sentences dealing with modification, and considers the possibility of automatic correction and processing of these structures.

(11)【公開番号】特開 2001－219655
(43)【公開日】平成 13 年 8 月 14 日
(54)【発明の名称】熱転写記録媒体及び画像形成方法
(21)【出願番号】特願 2000－30516
(22)【出願日】平成１２年２月８日
(72)【発明者】【氏名】椎名 義明
(57)【要約】【課題】インキの滲みによる解像力の低下を防ぐ事、また、ワックスを使用していると転写した画像を手で擦ったような場合の耐久性が足りず、画像が取れて無くなり易いことなどの点を改善して耐久性を増す事、そして特に、感熱転写の際の感熱転写シート基材の剥離時における熱転写記録層の箔切れ性（膜切れ性）が良く、且つ転写画像の光学濃度も高い事、これらを同時に充分達成することのできる熱転写記録媒体（感熱転写リボン）を提供する。【解決手段】支持体２上に少なくとも着色顔料と有機樹脂バインダーと無色又は淡色の微粒子とを主成分とする組成物から形成された熱転写記録層３が設けられた熱転写記録媒体１において該熱転写記録層が膜厚０．５～１．０μｍの範囲にあり、前記有機樹脂バインダーが平均分子量１００００～２００００の範囲の塩化ビニル－酢酸ビニル共重合樹脂である。

Fig.1 An example of patent form in Japanese [1]

(11) Publication number: **2001-219655**
(43) Date of publication of application: **14.08.2001**
(54) **THERMAL TRANSFER RECORDING MEDIUM AND IMAGE FORMING METHOD**
(21) Application number: **2000-030516**
(22) Date of filing: **08.02.2000**
(72) Inventor: **SHIBUYA KAZUMICHI, NAITO AKIRA, SHIINA YOSHIAKI**
(57) Abstract: PROBLEM TO BE SOLVED: To provide a thermal transfer recording medium (a heat-sensitive transfer ribbon) sufficiently and simultaneously achieving the prevention of lowering of resolution generated by the bleeding of an ink, the improvement of the points such as the insufficient durability, the tendency in which an image is easily removed and eliminated when a wax is used and the transferred image is rubbed by a hand and other points to increase the durability and particularly providing the good foil cutting properties (film cutting properties) when a heat-sensitive transfer sheet base is released in the heat-sensitive transfer and also providing the high optical density of the transferred image.

SOLUTION: The thermal transfer recording medium 1 is provided with a thermal transfer recording layer 3 formed of a composition composed at least of a color pigment, an organic resin binder and colorless or light-color fine particles as main components and formed on a substrate 2, and the film thickness of the thermal transfer recording layer is in the range of 0.5-1.0 μm, and the organic resin binder is a vinyl chloride-vinyl acetate copolymer resin of the average molecular weight in the range of 10,000-20,000….

Fig.2 An example of patent form translated in English from the example in Fig.1. [2]

2 Procedure

First, patent sentences are analyzed using the Japanese modification analyzer *Cabocha* [3]. Then, if there are mistakes, the resulting modification relationships are corrected by humans. The procedure is as follows:

(1) Patent data are randomly collected from the patent database [1].
(2) Comparing the result of modification analysis obtained from *Cabocha* with that obtained from humans, errors are found.
(3) Sub-sentences around the error are extracted, and are reanalyzed with *Cabocha*. The result should reveal the characteristics of the error.
(4) Classification of the results is performed.
(5) Consideration is given to the possibility of automatic processing and/or automatic correction (the algorithm structure will be completed in the near future, and is thus not included in this paper).

About 200 patent sentences extracted mainly from abstracts (problem to be solved and/or solution) are analyzed and corrected [4]. Classification is described in the next section.

3 Result of Classification

In the following classification, the example phrases and sentences written in Japanese are attached with English translation by human. The analyzed phrases and sentences for morphemes in the Figures have literal English translations in parentheses. The analysis of the modification is indicated by '< >' along with its line number and that of its modified form. The corrected results are indicated by '<< >>' and underlined.

3.1 Parallel Structure

Parallel structure can be divided into subcategories.

(a) Parallel structure of Japanese nouns dominated by the particle *to*

(S1) 生ごみを攪拌する攪拌手段と、処理槽内の生ごみ処理材及び生ごみを加熱する加熱手段と、(A stirring means to be established in a processing tub and to stir garbage – disposal material and a kitchen garbage, It has a heating means to heat the garbage – disposal material and kitchen garbage in a processing tub …)

(0) 生ごみを<係り受け 自="0" 先="1D"> (garbage)
(1) 攪拌する<係り受け 自="1" 先="2D"> (to stir)
(2) 攪拌手段と、<係り受け 自="2" 先="5D"> **《自="2" 先="７D"》** (stirring means)
(3) 処理槽内の<係り受け 自="3" 先="5D"> (processing tub)
(4) 生ごみ処理材及び<係り受け 自="4" 先="5D"> (disposal material)
(5) 生ごみを<係り受け 自="5" 先="6D"> (garbage)
(6) 加熱する<係り受け 自="6" 先="7D"> (heating)
(7) 加熱手段と、<係り受け 自="7" 先="なし"> (heating means)

Fig.2 The result of modification analysis in (S1)

The error in Fig.2 shows the incorrect analysis of the parallel relationship. The analyzer processed that there is a parallel relationship between the "stirring means" and "garbage"; this occurred due to a misunderstanding of the Japanese particle *wo* in line 5. In fact, it is "stirring means" and "heating means" that have the parallel relationship. Similarly, the relation of particles *ga*, *ni*, and *no* can also cause misunderstanding.

(b) Parallel structure involving Japanese commas

(S2) ベニコウジ色素の光退色防止方法、退色効果に優れたベニコウジ色素による食品、化粧品、医薬品の染色方法、(the dyeing approach of the photo-fading prevention approach of Monascus coloring matter, the food by Monascus coloring matter excellent in the fading effectiveness, cosmetics, and drugs,)

(0) ベニコウジ色素の<係り受け 自="0" 先="1D"> (Monascus coloring)
(1) 光退色防止方法、<係り受け 自="1" 先="5D"> **《自="1" 先="８D"》** (photofading prevention)
(2) 退色効果に<係り受け 自="2" 先="3D"> (in the fading effectiveness)
(3) 優れた<係り受け 自="3" 先="4D"> (excellent)
(4) ベニコウジ色素による<係り受け 自="4" 先="5D"> (by Monascus coloring)
(5) 食品、<係り受け 自="5" 先="6D"> (food)
(6) 化粧品、<係り受け 自="6" 先="7D"> (cosmetics)
(7) 医薬品の<係り受け 自="7" 先="8D"> (drugs)
(8) 染色方法、<係り受け 自="8" 先="なし"> (the dyeing approach)

Fig.3 The result of modification analysis of (S2)

In this sentence, the error is that “photofading prevention” and “food” have a parallel relationship. In fact, “photofading prevention” is parallel to “dyeing approach.”

3.2 Agreement between noun and verb

This class focuses on the modification between noun and verb in sentences with long modification phrases and/or clauses, or modification of the words in the modifying clause by another modifying clause. In some cases, the correct modification can be strictly defined. In other cases, however, even a human cannot define only one correct relationship, as another possibility exists.

（S3）支持部材が、内側空間の設けられた本体部分と、この本体部分の端部に設けられた平板状部とを備えており、(supporter material is equipped with the body part, in which inside space was established, and the plate-like section prepared in the edge of this body part,)

(0) 支持部材が、<係り受け 自="0" 先="7D">**《自="0" 先="９D"》** (supporter material)
(1) 内側空間の<係り受け 自="1" 先="2D"> (inside space)
(2) 設けられた<係り受け 自="2" 先="3D"> (established)
(3) 本体部分と、<係り受け 自="3" 先="6D"> (body part)
(4) この<係り受け 自="4" 先="5D"> (this)
(5) 本体部分の<係り受け 自="5" 先="6D"> (body part)
(6) 端部に<係り受け 自="6" 先="7D"> (in the edge of)
(7) 設けられた<係り受け 自="7" 先="8D"> (prepared)
(8) 平板状部とを<係り受け 自="8" 先="9D"> (the plate-like section)
(9) 備えており、<係り受け 自="9" 先="なし"> (equipped)

Fig.4 The result of modification analysis in (S3)

In Fig.4, the modified form of “supporter material” is analyzed as “prepared”, where “prepared” correctly modifies “the plate-like section”. In fact, “supporter material” and “equipped” have a modifying relationship.

3.3 Modification between subordinate clauses

This classification includes the relationship between verb subordinate clauses, and the relationship with adverbs and/or functional words.

(S4) メレンゲ混入のスフレ生地を入れ、メレンゲ混入のスフレ生地を入れた耐熱容器のスフレ生地より上側の内面に無塩バターを塗布し (dough is placed in the vessel, and the inner surface of the vessel above the dough in the vessel is coated with unsalted butter)

(0) メレンゲ混入の<係り受け 自="0" 先="1D">
(1) スフレ生地を<係り受け 自="1" 先="2D"> (dough)
(2) 入れ、<係り受け 自="2" 先="5D">**《自="2" 先="１１D"》** (is placed)
(3) メレンゲ混入の<係り受け 自="3" 先="4D">
(4) スフレ生地を<係り受け 自="4" 先="5D"> (dough)
(5) 入れた<係り受け 自="5" 先="6D"> (in the)
(6) 耐熱容器の<係り受け 自="6" 先="7D"> (vessel)
(7) スフレ生地より<係り受け 自="7" 先="11D"> (dough)
(8) 上側の<係り受け 自="8" 先="9D"> (above)
(9) 内面に<係り受け 自="9" 先="11D"> (the inner surface)
(10) 無塩バターを<係り受け 自="10" 先="11D"> (saltless butter)
(11) 塗布し、<係り受け 自="11" 先="なし"> (is coated)

Fig.5 The result of modification analysis in (S4)

In Fig.5, the analyzer interpreted that “is placed” modifies “in the”, which in turn adverbially modifies “vessel.” Correctly, “is placed” and “is coated” comprise the modifying relationship.

3.4 Modification noun clauses

Some mistakes are caused by the misanalysis of modifying noun clauses.

(S5) 製造工程等から排出されるトリメチルガリウム等の有害成分を含むガスを、(a gas containing noxious compounds such as trimethylgallium discharged from a manufacturing process)

(0) 製造工程等から<係り受け 自="0" 先="1D"> (from a manufacturing process)
(1) 排出される<係り受け 自="1" 先="2D">**《自="1" 先="５D"》** (discharged)
(2) トリメチルガリウム等の<係り受け 自="2" 先="3D"> (such as trimethylgallium)
(3) 有害成分を<係り受け 自="3" 先="4D"> (noxious compounds)
(4) 含む<係り受け 自="4" 先="5D"> (containing)
(5) ガスを、<係り受け 自="5" 先="なし"> (a gas)

Fig.6 The result of modification analysis in (S5)

In Fig.6, “discharged” can modify both “such as trimethylgallium” and “a gas.”

3.5 Conjunctives

Certain Japanese conjunctions have ambiguous modification potential, such as *matawa* (or), *oyobi* (and), and *narabini* (and).

(S6) 給水管が接続される給水継手及び給湯管が接続される給湯継手を備えるブラケットを(a bracket equipped with the hot-water supply joint to which the water supply joint and hot-water pipe to which a feed pipe is connected)

(0) 給水管が<係り受け 自="0" 先="1D"> (a feed pipe)
(1) 接続される<係り受け 自="1" 先="3D"> (is connected)
(2) 給水継手及び<係り受け 自="2" 先="3D">《自="2" 先="５D"》(supply joint and)
(3) 給湯管が<係り受け 自="3" 先="4D"> (hot-water pipe)
(4) 接続される<係り受け 自="4" 先="5D"> (is connected)
(5) 給湯継手を<係り受け 自="5" 先="6D"> (hot-water supply joint)
(6) 備える<係り受け 自="6" 先="7D"> (equipped)
(7) ブラケットを<係り受け 自="7" 先="なし"> (a bracket)

Fig.7 The result of modification analysis in (S6)

In Fig.7, “supply joint and” is analyzed as being modified by “a feed pipe.” This is caused by misunderstanding the purpose of the particle *ga* in lines 1 and 3. Actually, “supply joint” in line 2 and “hot-water supply joint” in line 5 have the parallel relationship, as shown in Fig.7.

3.6 Proper Representation in Patent Sentences

In patent sentences, there are typical representations about which the analyzer makes mistakes during normal processing.

(S7) 本発明は、腕時計のガラス面に半導体材料層を形成し、その半導体材料層上に、半導体デバイスの薄膜形成プロセスにより、温度センサーとして機能するトランジスタと、そのトランジスタの導通状態を検出することで温度を検出する温度検出回路と、温度検出回路によって検出された温度を温度データとして記憶する半導体メモリーとを形成してなる腕時計型温度センサーである。

(This invention is a sensor whenever [wrist watch mold temperature / which forms a semiconductor material layer in the glass side of a wrist watch, and comes to form the temperature detector which detects temperature according to the thin film formation process of a semiconductor device by detecting the switch-on of the transistor which functions as a thermo sensor, and its transistor, and the semiconductor memory which memorizes the temperature detected by the temperature detector as temperature data on the semiconductor material layer].)

(0) 本発明は、<係り受け 自="0" 先="24D">《自="0" 先="２８D"》(This invention is)
(1) 腕時計の<係り受け 自="1" 先="2D"> (on the wrist watch)
(2) ガラス面に<係り受け 自="2" 先="4D"> (glass surface)
(3) 半導体材料層を<係り受け 自="3" 先="4D"> (semiconductor material layer)
…
(22) 温度を<係り受け 自="22" 先="24D"> (temperature)
(23) 温度データとして<係り受け 自="23" 先="24D"> (as temperature data)
(24) 記憶する<係り受け 自="24" 先="25D"> (memorized)
(25) 半導体メモリーとを<係り受け 自="25" 先="26D"> (semiconductor memory)
(26) 形成して<係り受け 自="26" 先="27D"> (form)
(27) なる<係り受け 自="27" 先="28D">
(28) 腕時計型温度センサーである。<係り受け 自="28" 先="なし"> (wrist watch type temperature sensor)

Fig.8 The result of modification analysis in (S7)

In the above sentence, “This invention is” is a typical patent sentence.

In Fig.8, the initial phrase “The invention is” actually modifies the last phrase “wrist watch type temperature sensor.” Similarly, phrases such as “This paper is” and “This experiment is” are often used at the beginning of a sentence.

There are other typical phrases in this classification, such as noun phrases, and certain adverbial phrases employing the Japanese *deatte* (such as).

4 Concluding Remarks

This paper describes only the classification of modification error types in patent sentences. The number of analyzed sentences described above is only about 200, and as such, statistical analysis of sentence types and trends is not yet possible. A solution is being developed, which finds and automatically corrects these errors, but many more sentences must first be analyzed and classified.

If we are able to classify, detect, and modify the modification structure of these sentences

automatically, it will be a significant contribution to the improvement in quality of automatic patent translation, because of corrected modifying structures. In addition, the same method could be applied not only to patent sentences, but also to other types of structurally complicated Japanese sentences.

5 Acknowledgements

Mr. OSHIO Tadaaki, Japio, taught us the search and extraction methodology for patent data in both Japanese and English. This paper is based on discussion in the Special Interest Group in AAMT/Japio Research Committee.

References

[1] Industrial Property National Library in Japanese: http://www.ipdl.ncipi.go.jp/ homepg.ipdl

[2] Industrial Property National Library in English:
http://www.ipdl.ncipi.go.jp/ homepg_e.ipdl

[3] Modification Analyzer: "*Cabocha"*, Nara Institute of Science and Technology

[4] KANEDA Yuya: Analysis and Classification of Modification Relations in Patent Sentences (in Japanese), Bachelor Thesis of Yamagata University (2005)