

変更極小性を考慮した改正後法令文の機械翻訳

山腰貴大^{1,3} 小川泰弘² 外山勝彦¹

¹ 名古屋大学大学院情報学研究科 ² 名古屋市立大学データサイエンス学部

³ リーガル AI 株式会社

yamakoshi@legalai.jp ogawa@ds.nagoya-cu.ac.jp

toyama@is.nagoya-u.ac.jp

概要

一部改正に伴い既存の法令文を翻訳し直す際には、改正前の訳文の表現を維持し、改正により変化した部分のみを翻訳対象にすることが望ましい。この要件を変更極小性と呼び、改正後法令文の翻訳においては、流暢性、妥当性に並ぶ要件である。本研究では、変更極小性を考慮した機械翻訳の実現を目指し、改正前後の法令文のコーパスの構築、Transformer ベースの機械翻訳手法の開発、変更極小性の評価指標の開発を行った。また、改正後法令文の翻訳実験により、提案手法の有効性を検証した。

1 はじめに

法令は、一度制定すれば終わりではなく、社会、経済、自然環境等の変化に応じて適宜改正される。例えば、2023年にわが国において制定された法律は89件であるが、そのうち70件は、もっぱら既存の法律の一部を改正する一部改正法律である。

また、国際間取引等のトラブルを防ぐために、法令の翻訳を提供することが重要であるが[1]、法令改正の際には、その内容を国際的に周知するために、翻訳に速やかに反映することが望ましい。わが国では、法務省が日本法令外国語訳データベースシステム JLT (Japanese Law Translation Database System)¹⁾によって主要法令の日本語・英語対訳をウェブ上で提供しているが、時間や人手等の不足により、法令改正のペースに英訳の修正が追いついていない。

機械翻訳は、この問題を解決できる技術の一つである。昨今の技術発展により、人手による翻訳に近い品質の訳文を機械翻訳によって迅速に得られるようになった。しかし、改正によって一部修正された法令文を翻訳する際には、通常とは異なる注意を払う必要がある。具体的には、改正によって変化しな

い箇所の訳は変えないということである。そのような箇所の訳を不用意に変更すると、当該箇所も改正されたことと訳文の読者に受け取られるおそれがある。この要件を変更極小性 (focality) と呼び、改正後法令文の翻訳においては、通常の機械翻訳の要件である流暢性、妥当性と並ぶ第三の要件である。変更極小性は、法令に限らず、文の一部を改正することによりアップデートする文書、例えば、利用規約や製品マニュアルなどの翻訳でも重要となる要件である。

変更極小性を確保した改正後法令文の機械翻訳の実現のため、(1) コーパス、(2) 機械翻訳手法、(3) 評価指標を開発した。(1)については、変更極小性を考慮した改正後法令文コーパスを手で構築した。(2)については、翻訳対象の文(改正後の原文)だけでなく、改正前の訳文も入力し、後者の表現をコピーできる機構を備える Transformer [2] ベースの機械翻訳手法を開発した。(3)については、変更極小性を測定できる評価指標 ISDIT (Inclusive Score for Differential Translation) を開発した。

2 背景

法令の一部改正について 2.1 節で述べ、本研究で扱うタスクとその要件を 2.2 節で整理する。

2.1 法令の一部改正

わが国において、法令の一部改正を行う際は、いわば「ソフトウェアパッチ」の要領で被改正法令を書き換える[3]。すなわち、文字列や条項を置換、挿入、削除する指示(改め文)を一部改正法令において定め、改め文を改正対象に対して適用することにより法令を改正する。改め文の例を図 1 に示す。改め文によって規定される指示は、次のように整理できる[4]。

1. 文の一部に対する (a) 置換, (b) 挿入, (c) 削除

1) <https://www.japaneselawtranslation.go.jp/ja>

第百六十四条^①第四項を削り、^②第三項後段を削り、^③同項第一号中「の父母」を「(十五歳以上のものに限る。)」に改め、^④同項第二号中「前号に掲げる」を「…に対し親権を行う」に改め、^⑤同項第三号を削り、^⑥同項を同条第六項とし、^⑦同項の次に次の一項を加える。
7 特別養子適格の… (省略)

民法等の一部を改正する法律 (令和元年法律第34号) から抜粋

図1 改め文

新	旧
(省略)	(省略)
6 家庭裁判所は、特別養子縁組の成立の審判をする場合には、次に掲げる者の陳述を聴かなければならない。 ^②	3 家庭裁判所は、特別養子縁組の成立の審判をする場合には、次に掲げる者の陳述を聴かなければならない。この場合において、第一号に掲げる者の同意がないにもかかわらずその審判をするときは、その者の陳述の聴取は、審問の期日においてしなければならない。
③ 一 養子となる者 (十五歳以上のものに限る。)	一 養子となる者の父母
④ 二 養子となるべき者に対し親権を行う者 (養子となるべき者の父母及び養子となるべき者の親権者に対し親権を行う者を除く。) ^⑤ 及び養子となるべき者の未成年後見人	二 養子となるべき者に対し親権を行う者 (前号に掲げる者を除く。) 及び養子となるべき者の未成年後見人
(削除) ⑤	三 養子となるべき者の父母に対し親権を行う者及び養子となるべき者の父母の後見人
(削除) ⑦	4 家庭裁判所は、…
7 特別養子適格の… ^⑦	(省略)
(省略)	(省略)

図2 図1の改め文に対する新旧対照表

- 条, 項, 号等の法令構造に対する (a) 置換, (b) 挿入, (c) 削除.
- 法令構造に付与された番号の (a) 変更, (b) 割当, (c) シフト.
- 法令構造の番号と見出しの変更

一部改正の際には、改正法令の適用前後を比較できる新旧対照表が作成される。図2は、図1の改め文に対する新旧対照表の例である。改め文により変更を指示された箇所は、下線を付して表される。

2.2 改正後法令文の翻訳

本研究は、前節の分類1. (文の一部に対する修正) に伴う翻訳を扱う。以降、改正前の原文を旧原文、その訳文を旧訳文、改正後の原文を新原文、その訳文を新訳文といい、旧原文、旧訳文、新原文、新訳文からなる組を四つ組という。本研究のタスクは、既存の旧原文と旧訳文を参考に、新原文から新訳文を翻訳するタスクである。その際、新訳文に求める要件は、流暢性 (訳文が自然な言い回しであること) と妥当性 (原文の意味が訳文に過不足なく含まれていること) だけではない。その例を表1に示す。

表1に示す二つの新訳文#1と#2は、どちらも、言い回しは自然であり、新原文の意味を反映している。すなわち、どちらも流暢かつ妥当である。しかし、新訳文#2では、旧訳文中の“a document stating”を“writing describing”に変えているが、これは改正の対象ではない。そのため、日本語の分からない読者が旧訳文と新訳文#2を読んだとき、上述の変更も改正によるものだとして誤解するおそれがある。この点で、新訳文#2は新訳文として適切ではない。した

がって、本タスクにおいては、流暢性、妥当性に加えて、「改正の影響を受けていない部分の翻訳を不必要に変更しない」という要件も必要となる。本研究では、この要件を変更極小性 (focality) と呼ぶ。

3 関連研究

機械翻訳において変更極小性を高めるための方法に、原文と関連する文 (以下「関連原文」という。) の訳文 (以下「関連訳文」という。) 中の単語をコピーする方法がある。統計的機械翻訳 (SMT) でこの方法を導入した手法として、Koehn らの手法 [5] や、それを改正後法令文の翻訳に特化させた小酒井らの手法 [6] がある。これらの手法は、原文と関連原文の間の不変化部分を特定し、その不変化部分と対応する関連訳文中の翻訳を固定し、その他の部分のみを SMT によって新たに翻訳する。

ニューラル機械翻訳 (NMT) にこの方法を導入した手法として、Junczys-Dowmunt らの手法 [7] やそれを改良した Huang らの手法 [8] などがある。これらは、Transformer [2] をベースとし、原文に加えて関連訳文を入力する。訳文は、単語ずつ生成していくが、その際に、語彙だけでなく、関連訳文中の単語からも選択する。具体的には、関連訳文中の各単語について「コピーの尤度」を求め、語彙中の各単語の「生成の尤度」と足し合わせることで、最終的な尤度を計算する。

改正後法令文翻訳のための言語資源として、小酒井らが構築した四つ組データがある [6]。この四つ組データは、JLT の法令対訳データのうち、改正に伴う複数の版が存在する法令 17 件から構築され、四つ組 158 組と改正部分の情報を含む。しかし、訓練データとして用いるには、四つ組の量が極めて少ない。また、改正と無関係の変更を含む新訳文が存在し、変更極小性は完全に保たれていない [9]。

4 提案手法

本研究では、変更極小性を考慮した改正後法令文翻訳のために、コーパス、機械翻訳手法、評価指標を開発した。本節では、それぞれについて述べる。

4.1 コーパス

JLT は、改正に伴い版の異なる法令対訳データを提供しており、それをを用いて四つ組を構築することはできるものの、版の間の変更極小性は考慮されていない。そこで、JLT のデータを旧原文・旧訳文、

表 1 新訳文の比較（実線の下線が改正箇所）

種類	文
旧原文	前項の申立ては、海難の事実を示して、書面でこれをしなければならない。
新原文	前項の申立ては、海難の事実及び受審人に係る職務上の故意又は過失の内容を示して、書面でこれをしなければならない。
旧訳文	The request shall be made in a document stating the facts of the marine accident.
新訳文#1	The request shall be made in a document stating the facts of the marine accident and the details of the intentional or negligent act committed in the course of duties of the examinee.
新訳文#2	The request shall be made in writing describing the facts of the marine accident and the details of the intentional or negligent act committed in the course of duties of the examinee.

e-Gov 法令検索のデータを新原文とし²⁾、旧訳文からの変更極小性を保った新訳文を人手で作成することにより、四つ組コーパスを構築した [9]。2023 年 11 月までに、148 改正 4,630 組の四つ組を構築し、今後も拡充する予定である。

4.2 機械翻訳手法

流畅性、妥当性、変更極小性の鼎立を目指した機械翻訳手法として、Copiable Translation Transformer (CTT) を開発した [10]。CTT の構造を図 3 に示す。CTT は、旧訳文と新原文を入力とし、新訳文を生成する Transformer ベースの手法であり、旧訳文中の単語をコピーできる Copying Mechanism [11] を備えている。具体的には、単語出力の際に、語彙中の各単語の「生成」の尤度と、旧訳文中の各単語の「コピー」の尤度を求め、もっとも高い尤度の単語を出力する。その際、「生成」「コピー」の別の情報、また、「コピー」の場合は、旧訳文中の何番目の単語をコピーしたのかの情報を併せて出力する。

Copying Mechanism を備える手法に、3 節で述べた Junczys-Dowmunt らの手法 [7] やそれを改良した Huang らの手法 [8] などがある。これらの手法は、語彙中の各単語について、生成の尤度とコピーの尤度を合算したうえで最終的な尤度を求めている。そのため、出力文中の各単語からは、その単語が生成されたのかコピーされたのかを区別できない。一方、CTT は、出力文中の各単語について、「生成」「コピー」の別、後者の場合はコピー元の単語の場所も出力する。この情報を用い、旧訳文中においてコピーされなかった単語と新原文中において生成された単語をマーキングすることにより、新旧対照表の英訳を下線情報込みで自動生成できる。

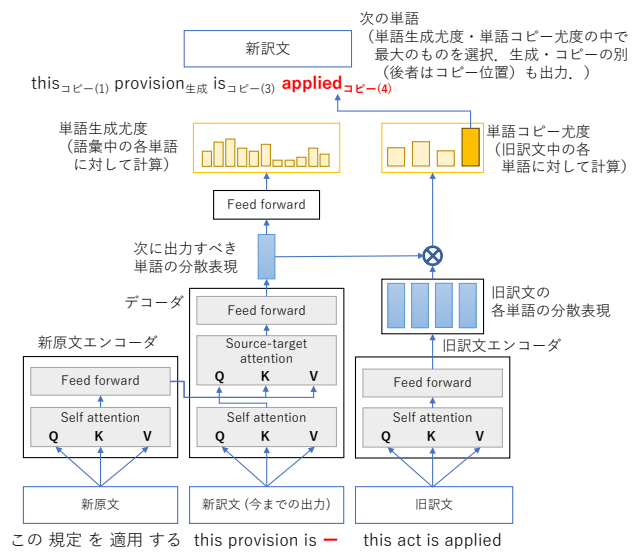


図 3 Copiable Translation Transformer の構造

4.3 評価指標

訳文の自動評価で広く用いられる BLEU [12] や RIBES [13] では、変更極小性は評価対象になっていない。そこで、変更極小性を評価するための指標 ISDIT (Inclusive Score for Differential Translation) を開発した [9]。ISDIT の式を以下に示す。

$$ISDIT(S_{出力文}; S_{旧訳文}, S_{参照訳}) = \text{Pre}(S_{出力文}; S_{参照訳}) \cdot \text{Rec}(S_{出力文}; S_{旧訳文}, S_{参照訳}) \quad (1)$$

ISDIT は、(1) 参照訳に対する出力文の n -gram 適合率 $\text{Pre}(S_{出力文}; S_{参照訳})$ と (2) 参照訳と旧訳文に共通して出現する n -gram に対する出力文の n -gram 再現率 $\text{Rec}(S_{出力文}; S_{旧訳文}, S_{参照訳})$ の積によって求める。すなわち、参照訳の表現が出力文においてより多く出現するほど、また、旧訳文に存在する表現を出力文がより多く拾えているほど高い評価となる。

5 実験

改正後法令文の翻訳実験について述べる。

2) e-Gov 法令検索 (<https://elaws.e-gov.go.jp/>) は、日本の法令を公開しているデジタル庁所管のウェブサイトである。同サイトは法令改正を速やかに反映する一方、JLT はそうではない。

表2 比較システム

名称	手法	学習データ
Trm _S	Transformer	四つ組コーパス
Trm _L	Transformer	対訳コーパス
Kozakai	小酒井らの手法	対訳コーパス
Huang	Huang らの手法	四つ組コーパス
CTT	CTT (提案手法)	四つ組コーパス

表3 実験結果

モデル	BLEU	RIBES	ISDIT
Trm _S	24.63	57.19	12.40
Trm _L	80.31	93.44	71.36
Kozakai	82.79	92.04	77.53
Huang	70.33	92.27	72.78
CTT	74.46	93.19	77.64

5.1 実験設定

機械翻訳手法として広く用いられている Transformer [2], 変更極小性に強い手法である小酒井らの手法 [6], Huang らの手法 [8] を提案手法と比較する。比較するシステムの一覧を表 2 に示す。表 2 中の「四つ組コーパス」は、4.1 節の四つ組コーパスであり、四つ組 4,630 組中 4,297 組を学習データとして用いた。なお、Huang 及び CTT の学習には、新原文、新訳文、旧訳文を用い、Trm_S の学習には、新原文と新訳文を用いた。表 2 中の「対訳コーパス」は、JLT から取得した対訳文 391,758 組から構築し、すべて学習データとして用いた。テストデータとして、上述の四つ組 4,630 組から学習データとは別に 201 組を選択して用いた。コーパスの規模による性能差を調査するため、Transformer は、Trm_S と Trm_L の 2 モデルを学習させた。

各手法は、BLEU, RIBES, ISDIT で評価した。学習時のハイパーパラメータは、付録 A に示す。

5.2 結果

実験結果を表 3 に示す。BLEU, RIBES, ISDIT において最大スコアを記録した手法は、それぞれ、Kozakai, Trm_L, CTT であった。Kozakai が BLEU, ISDIT で高いスコアを獲得したのは、改正の対象外と判断した旧訳文の表現を固定し、当該表現を出力文中で確実に反映できたためと思われる。一方、RIBES は語順の一致度により評価するため、流暢な文を出力できる NMT の手法 Trm_L が高いスコアを獲得したと考えられる。CTT は、規模の小さい四つ組コーパスを用いて学習したが、対訳コーパスを用いた Kozakai や Trm_L と並ぶ結果となった。

出力例を表 4 に示す。CTT, Kozakai は、改正内容である「前二項」から「前三項」への修正を反映し、その他の表現は旧訳文のものを維持できた。一方、Trm_L は、流暢かつ妥当な翻訳であるが、改正対象ではない「証明書」を“a certificate of identification”に変更してしまうなど、変更極小性が損なわれている。

表4 出力例 (実線の下線が改正箇所)

種類	文
旧原文	前二項の規定により立入検査をする職員は、その身分を示す証明書を携帯し、関係人にこれを提示しなければならない。
新原文	前三項の規定により立入検査をする職員は、その身分を示す証明書を携帯し、関係人にこれを提示しなければならない。
旧訳文	an officials conducting an on-site inspection pursuant to the provisions of the preceding two paragraphs must carry an identification card and present it to the persons concerned .
参照訳	an officials conducting an on-site inspection pursuant to the provisions of the preceding three paragraphs must carry an identification card and present it to the persons concerned .
Kozakai	an officials conducting an on-site inspection pursuant to the provisions of the preceding three paragraphs must carry an identification card and present it to the persons concerned .
Trm _L	officials conducting on-site inspections pursuant to the provisions of the preceding three paragraphs must carry a certificate of identification and present it to the persons concerned .
CTT	an officials conducting an on-site inspection pursuant to the provisions of the preceding three paragraphs must carry an identification card and present it to the persons concerned .

6 まとめ

本稿では、改正後法令文向けに開発した機械翻訳手法について述べた。改正後法令文の翻訳においては、流暢性、妥当性に加えて変更極小性も翻訳の要件となる。そこで、変更極小性を考慮した改正後法令文コーパス、機械翻訳手法 CTT, 評価指標 ISDIT を提案し、提案手法を実験により評価した。今後は、コーパスの拡充と CTT の改善に取り組む。コーパスの拡充は、NMT の学習データとしては規模が依然小さいため、自動生成によるコーパスの大幅拡充を検討する。CTT は、旧訳文をそのまま出力することが多く (付録 B), この問題の解決を目指す。

謝辞

本研究は JSPS 科研費 JP21H03772, JP22H03901 の助成を受けたものです。

参考文献

- [1] 外山勝彦, 小川泰弘. 自然言語処理の応用に基づく法令外国語訳支援. 人工知能学会誌, Vol. 23, No. 4, pp. 521–528, 2008.
- [2] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In **Proceedings of Advances in Neural Information Processing Systems 30**, pp. 6000–6010, 2017.
- [3] 法制執務研究会 (編). 新訂ワークブック法制執務第2版. ぎょうせい, 2018.
- [4] Yasuhiro Ogawa, Shintaro Inagaki, and Katsuhiko Toyama. Automatic consolidation of Japanese statutes based on formalization of amendment sentences. **New Frontiers in Artificial Intelligence: JSAI 2007 Conference and Workshops, Revised Selected Papers, Lecture Notes in Computer Science**, Vol. 4914, pp. 363–376, 2008.
- [5] Phillip Koehn and Jean Senellart. Convergence of translation memory and statistical machine translation. In **AMTA Workshop on MT Research and the Translation Industry**, pp. 21–31, 2010.
- [6] 小酒井款雄, 小川泰弘, 大野誠寛, 中村誠, 外山勝彦. 新旧対照表の利用による法令の英訳修正. 言語処理学会第23回年次大会論文集 (NLP2017), pp. 859–862, 2017.
- [7] Marcin Junczys-Dowmunt and Roman Grundkiewicz. MS-UEdin submission to the WMT2018 APE shared task: Dual-source transformer for automatic post-editing. In **Proceedings of the Third Conference on Machine Translation: Shared Task Papers**, pp. 822–826, 2018.
- [8] Xuancheng Huang, Yang Liu, Huanbo Luan, Jingfang Xu, and Maosong Sun. Learning to copy for automatic post-editing. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)**, pp. 6122–6132, 2019.
- [9] Takahiro Yamakoshi, Takahiro Komamizu, Yasuhiro Ogawa, and Katsuhiko Toyama. Evaluation scheme of focal translation for Japanese partially amended statutes. In **Proceedings of the 8th Workshop on Asian Translation (WAT2021)**, pp. 124–132, 2021.
- [10] Takahiro Yamakoshi, Yasuhiro Ogawa, and Katsuhiko Toyama. Differential-aware transformer for partially amended sentence translation. In Yasufumi Takama, Katsutoshi Yada, Ken Satoh, and Sachiyo Arai, editors, **New Frontiers in Artificial Intelligence, JSAI 2022 Conference and Workshops, Revised Selected Papers, Lecture Notes in Computer Science**, Vol. 13859, pp. 5–22, 2023.
- [11] Jiatao Gu, Zhengdong Lu, Hang Li, and Victor O.K. Li. Incorporating copying mechanism in sequence-to-sequence learning. In **Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics**, pp. 1631–1640, 2016.
- [12] Kishore Papineni, Salim Roukos, Todd Ward, and Wei Jing Zhu. BLEU: a method for automatic evaluation of machine translation. In **Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics**, pp. 311–318, 2002.
- [13] 平尾努, 磯崎秀樹, 須藤克仁, Kevin Duh, 塚田元, 永田昌明. 語順の相関に基づく機械翻訳の自動評価法. 自然言語処理, Vol. 21, No. 3, pp. 421–444, 2014.
- [14] Takahiro Yamakoshi, Takahiro Komamizu, Yasuhiro Ogawa, and Katsuhiko Toyama. Differential translation for Japanese partially amended statutory sentences. In Naoaki Okazaki, Katsutoshi Yada, Ken Satoh, and Koji Mineshima, editors, **New Frontiers in Artificial Intelligence. JSAI-isAI 2020. Lecture Notes in Computer Science**, Vol. 12758, pp. 162–178, 2021.

A 学習時のハイパーパラメータ

Trm_S, Trm_L 及び CTT の学習における主なハイパーパラメータは、次のとおりである。

- 隠れ層の数：6
- アテンションヘッド数：8
- 次元数：512
- バッチサイズ：8
- ドロップアウト率：0.1
- 最大系列長：256
- 反復数：20,000 (Trm_S), 2,000,000 (Trm_L)
- ビームサーチ数：4

Huang の学習におけるハイパーパラメータは、以下を除き、著者 GitHub リポジトリ³⁾が提示する設定を使用した。

- バッチサイズ：1,000 (実際のバッチサイズは、入力系列の長さに基づいた可変長であり、平均 15 程度である.)
- 反復数：20,000
- Evaluation step を実行しない
- Pre-training しない

なお、Trm_S, Trm_L, CTT, Huang は、Google Colab を用いて学習した。

B 提案手法の失敗例

提案手法において、旧訳文をそのまま出力した事例が多く見られた。表 5 にその例を示す。

表 5 の CTT#1 から CTT#4 は、ビームサーチによる CTT の 4-best 出力であり、CTT#1 がベスト出力 (すなわち、提案手法の出力) である。CTT#1 は、旧訳文そのままである一方、CTT#4 は、改正内容である「又は異議申立人」の削除を適切に反映した。また、Kozakai や Trm_L は、旧原文と新原文の間で実質的な変化がない⁴⁾「審査請求人」の訳を変えた点で変更極小性を達成できなかったものの、前述の改正内容を反映した文を出力した。

そのため、CTT の各出力候補を Kozakai や Trm_L 等の出力文と比較し、最適な出力候補を選択するアプローチ [14] を導入することにより、旧訳文をそのまま出力する問題の改善が期待できる。

3) <https://github.com/THUNLP-MT/L2Copy4APE>

4) 改め文は、文字レベルの単純な差分ではなく、一定の意味のまとまりを改正対象とする [3] ため、文字列として実質的な変化がない表現を改正対象に含みうる。

表 5 旧訳文をそのまま出力した例

種類	文
旧原文	経済産業大臣は、前条の意見の聴取の期日及び場所を定め、 <u>審査請求人又は異議申立人に通知しなければならない。</u>
新原文	経済産業大臣は、前条の意見の聴取の期日及び場所を定め、 <u>審査請求人に通知しなければならない。</u>
旧訳文	the minister of economy , trade and industry must specify the date and place of the hearing of opinions prescribed in the preceding article , and notify the requestor for review or <u>the petitioner for objection .</u>
参照訳	the minister of economy , trade and industry must specify the date and place of the hearing of opinions prescribed in the preceding article , and notify <u>the requestor for review .</u>
Kozakai	the minister of economy , trade and industry must specify the date and place of the hearing of opinions prescribed in the preceding article , and notify <u>the applicant for examination for the .</u>
Trm _L	the minister of economy , trade and industry must specify the date and place of the hearing of opinions prescribed in the preceding article , and notify <u>the applicant for examination of such date and place .</u>
CTT#1	the minister of economy , trade and industry must specify the date and place of the hearing of opinions prescribed in the preceding article , and notify <u>the requestor for review or the petitioner for objection .</u>
CTT#2	the minister of economy , trade and industry must specify the date and place of the hearing of opinions prescribed in the preceding <u>paragraph , and notify the requestor for review or the petitioner for objection .</u>
CTT#3	the minister of economy , trade and industry must specify the date and place of the hearing of opinions prescribed in the preceding article , and notify <u>the request the requestor for review or the petitioner for objection .</u>
CTT#4	the minister of economy , trade and industry must specify the date and place of the hearing of opinions prescribed in the preceding article , and notify <u>the requestor for review .</u>