機械翻訳もある総合的な翻訳サービスの模索
～金融・IR分野を例に～

開示書類翻訳におけるMTの可能性

宝印刷株式会社
グローバルリレーションズ部
三輪 哲也
2019年11月19日

©宝印刷（株） 無断複製を禁ず

宝印刷グループ

宝印刷グループは、制度開示・任意開示を含めた企業の情報開示における総合的な支援事業を中心として事業を展開しています

開示書類作成ツールの提供

コンサルティング

原稿企画・作成・翻訳

TAKARA グループ

研究部チェック

制作部チェック

印刷・納品

提出先 金融庁 EDINET

提出先 日本取引所 TDnet（適時開示情報伝達システム）

提出先 株主様

※EDINET･･･金融商品取引法に基づく有価証券報告書等の開示書類に関する電子開示システム。Electronic Disclosure for Investors' NETworkの略。
※TDnet ･･･東京証券取引所の運営する適時開示情報伝達 システム。Timely Disclosure networkの略。

©宝印刷（株）無断複製を禁ず

検証環境と検証エンジン

【使用自動翻訳プラットホーム】

みんなの自動翻訳@TexTra®

国立研究開発法人情報通信研究機構（NICT）が開発した自動翻訳サイト

【検証自動翻訳エンジン】

汎用NMT 【日本語 - 英語】	現在、みんなの自動翻訳@TexTra®で無償で提供されているエンジン
金融NMT 【日本語 - 英語】	金融関連のコーパスでアダプテーションをしたRNNモデルのエンジン
金融NT 【日本語 - 英語】	金融関連のコーパスでアダプテーションをしたTransformerモデルのエンジン

©宝印刷（株） 無断複製を禁ず

検証書類と検証方法

翻訳言語	検証数	検証訳文	書類	翻訳量（和文文字数）	
日本語 ↓ 英語	各書類5社 計：15社	各社・各書類　4バージョン ① 人間翻訳版 ② 汎用NMT版 ③ 金融NT版 ④ 金融NMT版 各書類20ファイル 計60ファイル	招集通知 狭義・参考書類	最大	22,700
				最小	9,600
				平均	12,680
			短信 定性情報	最大	5,000
				最小	1,800
				平均	3,460
			適時開示	最大	1,400
				最小	700
				平均	1,040

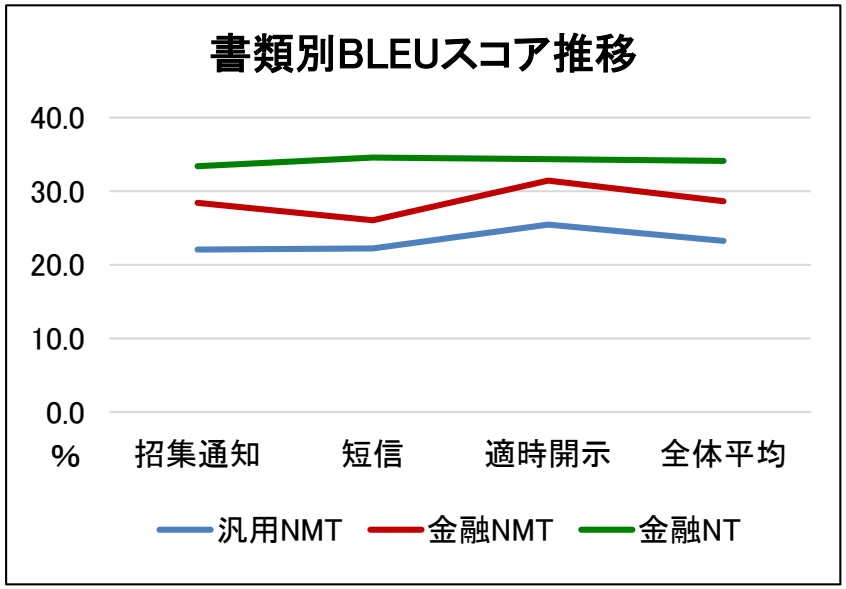
©宝印刷（株）無断複製を禁ず

検証書類と検証方法

評価方法	１．BLEUスコア ２．翻訳レビューワーによる評価
レビュー条件	上記①～④の訳文名をランダムにA～Dのファイル名を付与 レビューワーには、A～Dのどの訳文が①～④のバージョンであるかの情報は提供せず、レビューを依頼
レビューワー評価①	次の5つのカテゴリーで、A～Dの訳文をランク付け １）数値 ２） 固有名詞等の数値以外の定型用語 ３） 誤訳 ４） 訳抜け ５） 文章力
レビューワー評価②	各訳文について、気が付いた点を文章で評価

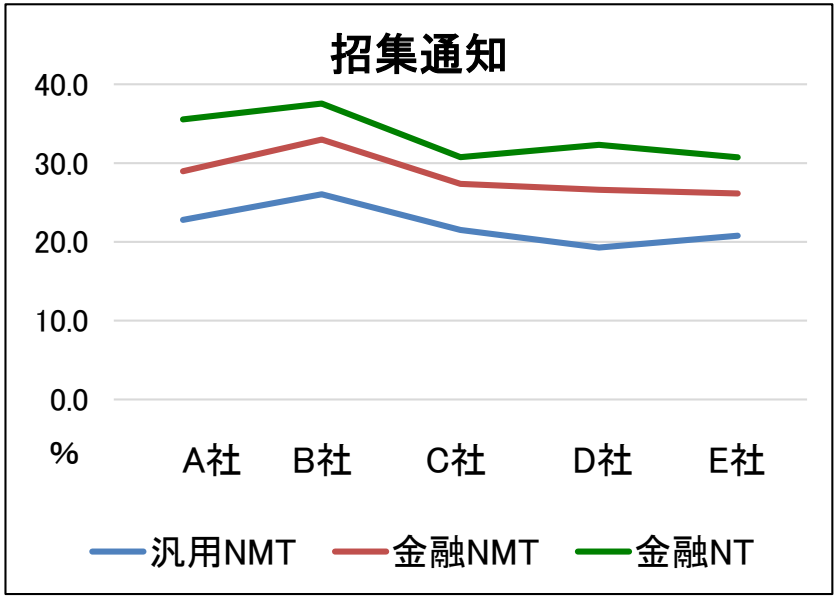
©宝印刷（株） 無断複製を禁ず

BLEUスコア



平均（%）	汎用NMT	金融NMT	金融NT
招集通知	22.1	28.4	33.4
短信	22.2	26.1	34.6
適時開示	25.5	31.4	34.4
全体平均	23.3	28.6	34.1

%

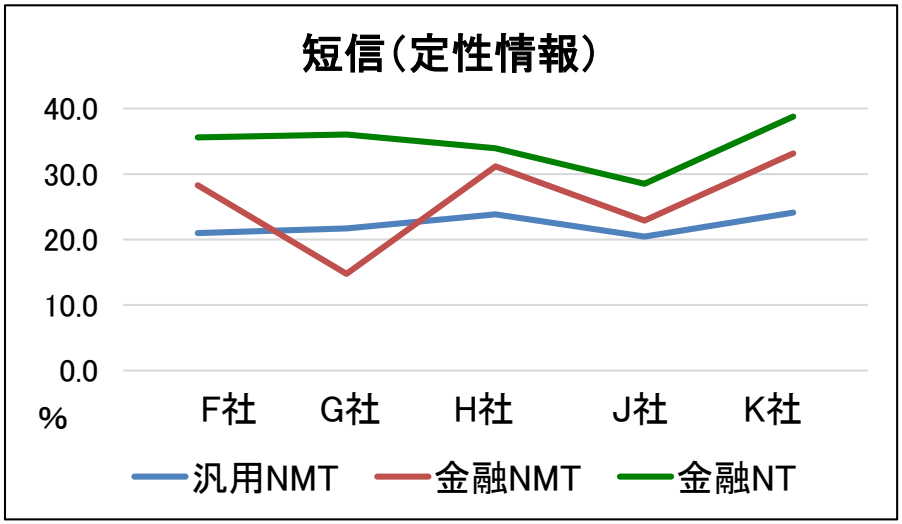


招集通知	汎用NMT	金融NMT	金融NT
A社	22.8	29.0	35.6
B社	26.0	33.0	37.6
C社	21.5	27.4	30.8
D社	19.3	26.6	32.3
E社	20.8	26.1	30.7

%

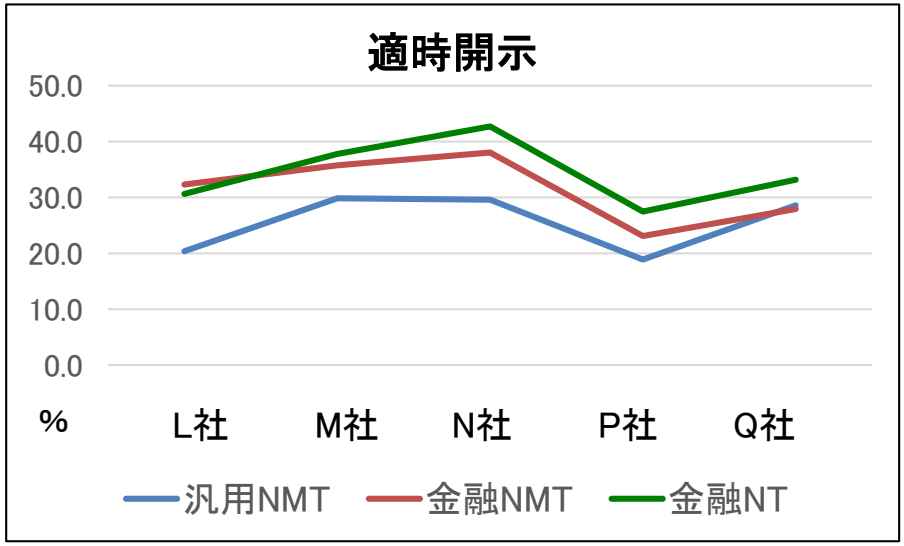
©宝印刷（株）無断複製を禁ず

BLEUスコア



短信	汎用NMT	金融NMT	金融NT
F社	21.0	28.3	35.6
G社	21.7	14.8	36.1
H社	23.9	31.2	34.0
J社	20.4	22.9	28.5
K社	24.1	33.2	38.8

%

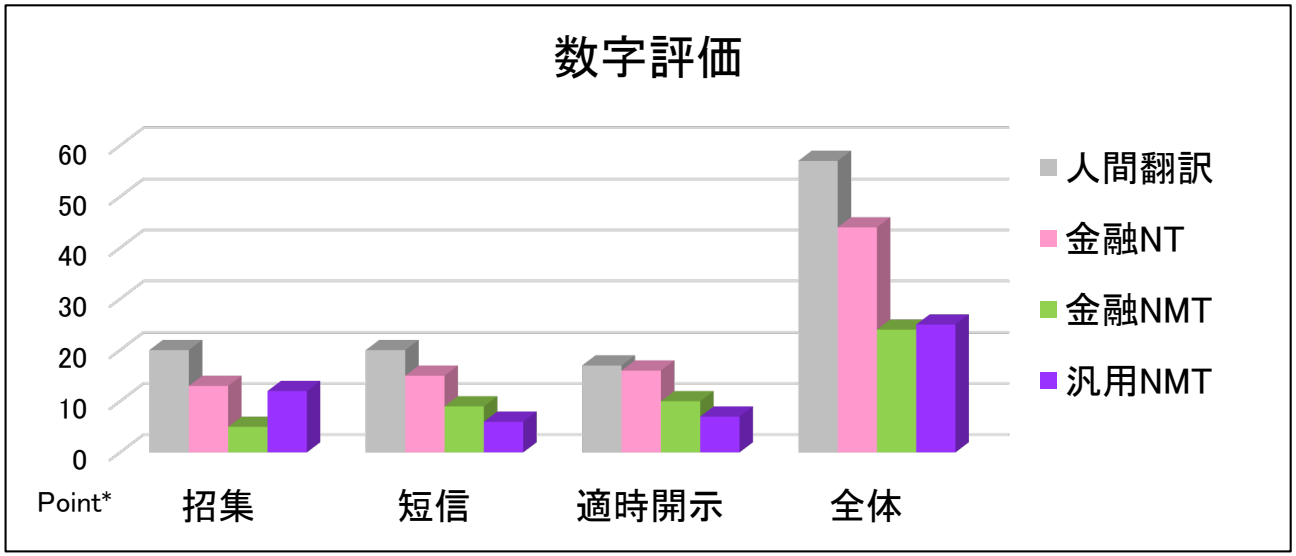


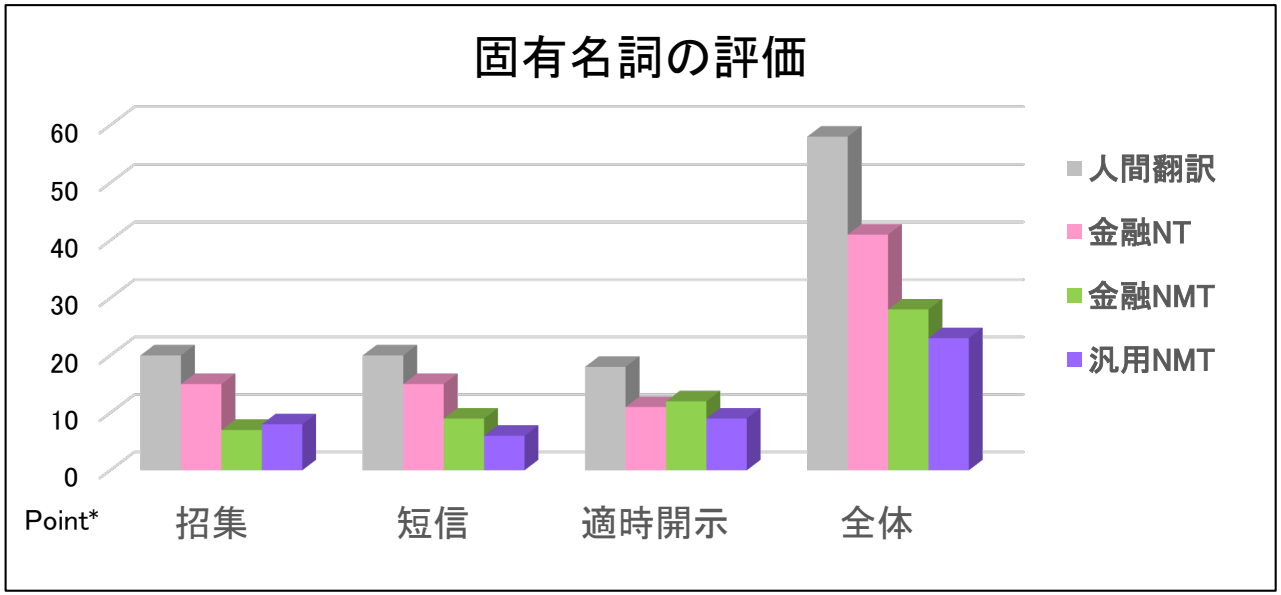
適時開示	汎用NMT	金融NMT	金融NT
L社	20.4	32.3	30.6
M社	29.8	35.8	37.8
N社	29.6	38.0	42.7
P社	18.9	23.1	27.5
Q社	28.6	27.9	33.2

%

©宝印刷（株）無断複製を禁ず

数字と固有名詞の評価





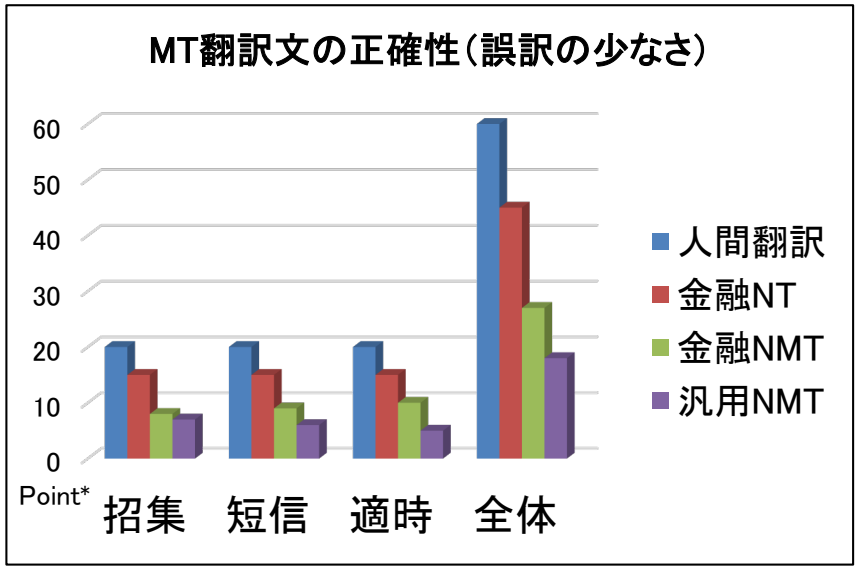
Point:
各順位になった件数
×
順位別係数

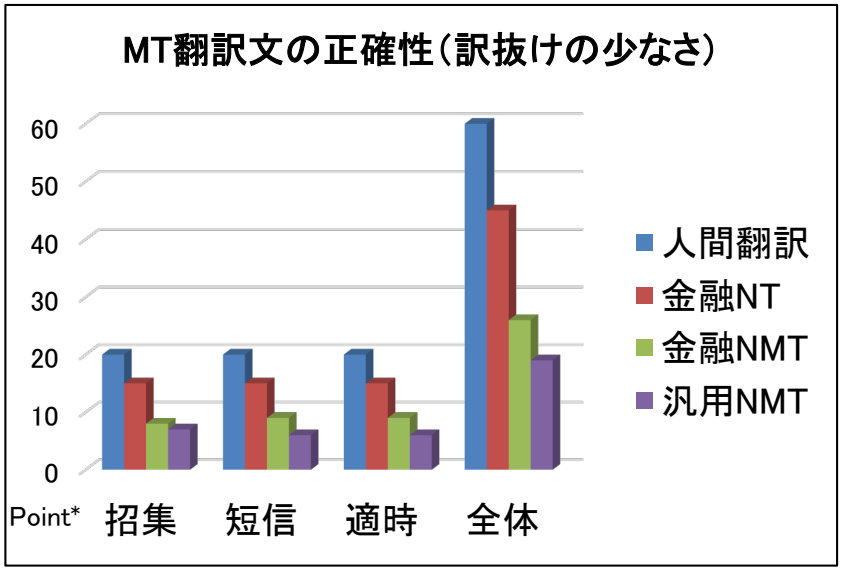
例:
1位の件数:2回
2位の件数:1回
3位の件数:2回
4位の件数:0回

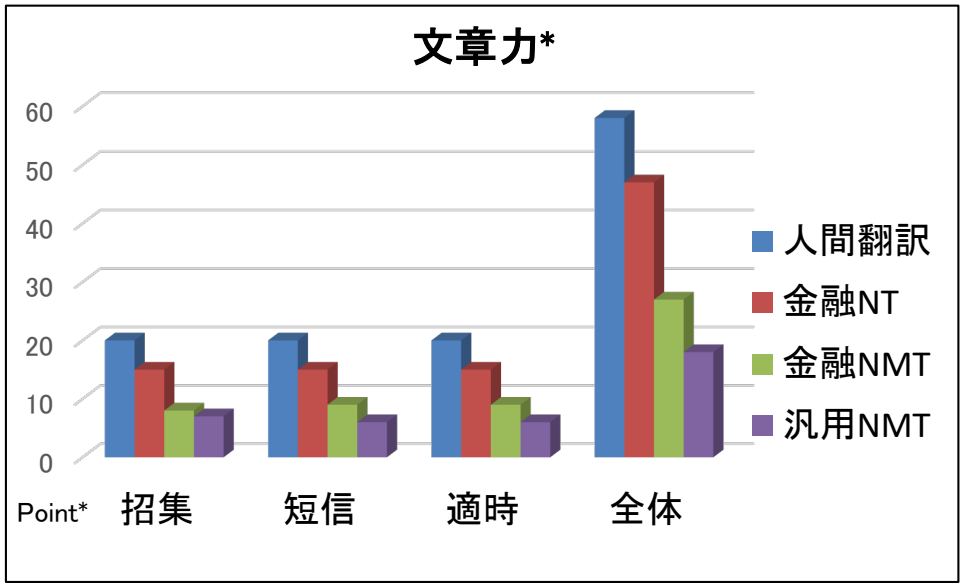
2回×4 = 8
1回×3 = 3
2回×2 = 4
計: 15 point

©宝印刷(株) 無断複製を禁ず

誤訳・訳抜け・文章力の評価







英文力:
英文としての自然さや構成力を指す

©宝印刷（株）無断複製を禁ず

MT活用の可能性

1. Secureな環境でのMT翻訳

インサイダー情報を含む書類 → 牢固なセキュリティが確保された環境

2. 数字翻訳の改善

和文と全く異なる数字を訳出

期間・日付の訳文が不正確

→ 正確な数字・期間・日付を訳出

3. 和文・英文セットでの標準化

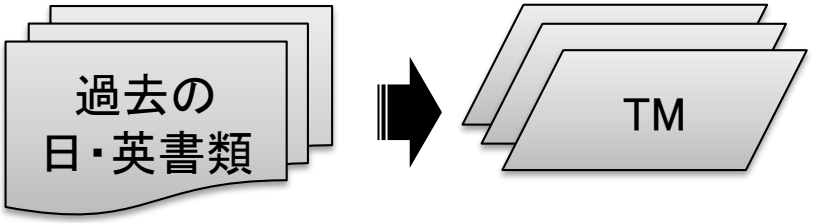
同じ内容に、複数の和文表現が存在 → 同じ内容を、同じ和文表現に

同じ和文に、複数の英文表現が存在 → 同じ和文は、同じ英文表現に

©宝印刷（株）無断複製を禁ず

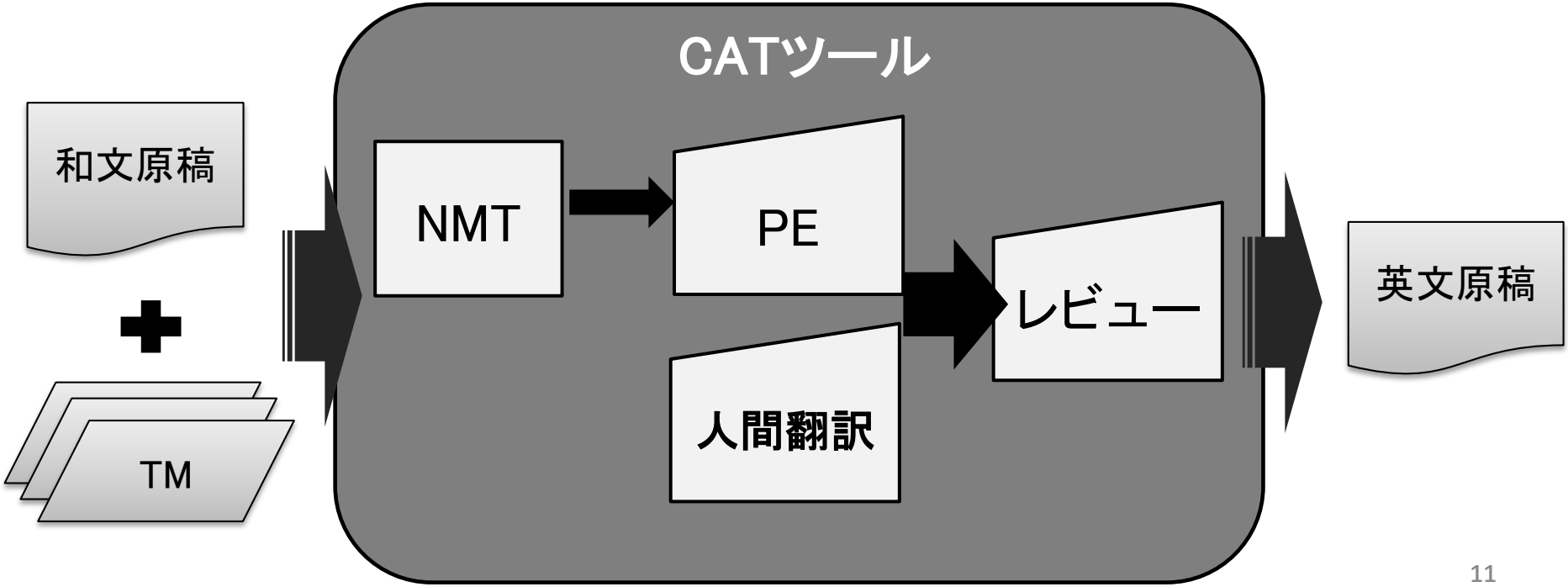
将来的な作業フロー

過去書類の有効活用



NMTの改善

NMTと人間翻訳の併用



©宝印刷（株）無断複製を禁ず

ご静聴ありがとうございました