パテントファミリーを用いた米国特許の構造解析および技術動向マップの自動作成

仲光純[1]　福田悟志[1]　難波英嗣[1]

概要：特許庁では，技術の発展が期待される分野や社会的に注目されている分野を対象に，特許出願技術動向調査を実施しており，これまでに 250 件以上の報告書と論文・特許リストが公表されている．しかし，これらの報告書は，時間の経過と共に古くなってしまう．そこで，特許出願技術動向調査報告書の自動更新を検討する．本研究では，報告書の自動更新に向けた第一歩として，米国特許の構造を分析することを目的とする．「効果」，「課題」，「手段」といった構造タグが付与された米国特許 4,410 件のデータセットを，構造タグが付与された日本国特許と，それに対応する米国特許（パテントファミリー）を用いて自動作成した．この手法の有効性を確認するために，いくつかの実験を行った．実験の結果，F 値 0.11 が得られた．

キーワード：特許，文書構造解析，機械翻訳，機械学習

Analyzing the Structure of U.S. Patents Using Patent Families for Automatic Creation of Technical Trend Maps

JUN NAKAMITSU[1]　SATOSHI FUKUDA[1]
HIDETSUGU NANBA[1]

Abstract: The Japan Patent Office carries out the patent application technical trend survey for fields in which the development of technologies is expected, or fields to which social attention is being paid, and more than 250 reports and their lists of research papers and patents have been released up to now. Unfortunately, these reports will become obsolete as time goes on. Therefore, we investigate automatic update of patent application technical trend survey reports. In this study, we aim to analyze the structure of U.S. patents as a first step toward the automatic update of the reports. We automatically created a dataset consisting of 4,410 U.S. patents with structure tags, such as “effect,” “task,” and “method,” by using Japanese patents, in which structure tags were manually assigned, and their corresponding U.S. patents (patent families). To confirm the effectiveness of our methods, we conducted some experiments. From the experimental results, we obtained F-measure of 0.11.

Keywords: patent, document structure analysis, machine translation, machine learning

1. はじめに

研究者や企業が新規開発を検討する際，最新の研究動向を把握するために，特許情報の活用が重要である．一方，世界中で発表される特許全てに目を通すことは困難である．このような状況で，特許庁が作成する特許出願技術動向調査報告書は，特定分野の技術動向を知る上で非常に有効である．しかし，この報告書の作成には非常にコストがかかるため，自動作成技術が望まれている．本研究では，特許出願技術動向調査報告書の自動作成を目的としており，その第一歩として，米国特許の構造解析を目指す．

特許出願技術動向調査報告書では，特定分野の特許や論文を，各研究分野における課題や技術等の情報を観点ごとにまとめた，技術分析軸と呼ばれる分類軸に分類することにより，特定分野の技術動向を俯瞰することを可能にしている．日本国特許は，「発明の効果」（以後，「効果」），「発明が解決しようとする課題」（以後，「課題」），「課題を解決するための手段」（以後，「手段」）等の項目に沿って記述されるため，日本国特許を技術分析軸に分類する手がかりとなる文は容易に見つかる．一方で，米国特許にはそのような項目は明示的には存在しないため，米国特許を対象にした分析は容易ではない．そこで，本研究ではパテントファミリーを利用する．パテントファミリーは，言語や構造は異なるが，出願書類を構成する文章は，ほぼ対訳となっている．そこで，本研究では，「効果」や「課題」などの構造タグが付与された日本国特許内の文の対訳文である米国特許内の文に同様の構造タグを付与することで，米国特許の構造を明らかにする．対訳文は，英語に翻訳した日本国特許の文とのコサイン類似度が最も高くなる米国特許内の文とする．さらに，構造化した米国特許を教師データとして機械学習を行うことで，米国特許から，「効果」，「課題」，「手段」のそれぞれに関する文を自動的に抽出するシステムを構築する．

本研究により，米国特許から技術動向に関する重要文を抽出することで，各特許を技術分析軸に分類する際の手がかりとなる．また，競合他社の現状技術を把握し，世の中の需要に見合った研究開発を効率的に行うことも可能であ

1 中央大学理工学研究科
Graduate School of Science and Engineering, Chuo University

る．加えて，特許庁が新たな技術分析軸の追加を査定する上で有効である．

2. 特許技術動向調査報告書

特許庁では，市場創出に関する技術分野，国の政策として推進すべき技術分野を中心に，今後の進展が予想される技術テーマを選定し，特許出願技術動向調査を実施して，報告書を作成している．報告書では，技術分析軸と呼ばれる，各研究分野における課題や技術等の情報を観点ごとにまとめた分類軸が定義されており，特定分野の技術動向を俯瞰することが可能となっている. 表1に平成29年度の「自動走行システムの運転制御」における技術分析軸の例を示す．表1からわかるとおり，技術分析軸の一つに，運転支援システムという大きな技術課題があり，その中に運転負荷軽減システム，さらにその中に車線維持支援や，駐車支援など、より詳細な技術がある．このトピックには，分析軸が243個定義されている．

また，報告書における本文の他に，報告書を作成するために収集された各テーマに関する国内外の論文と特許のリストがExcelファイルとして作成されている．このリストでは，各論文と特許が，どの技術分析軸に属するか記載されている．表1に自動走行システムの運転制御に関する技術分析軸に分類された特許の例を示す. また, 各特許には，表2に示すような英数字の羅列である固有のIDが与えられている．表1と表2を照らし合わせると，特許JP-A-2011-141747は，右左折支援という技術分析軸に属していることが分かる．

表1　自動走行システムの運転制御に関する技術分析軸の一部

運転支援システム	運転負荷軽減システム	車線維持支援(LKAS)		D11
		定速走行・車間距離制御(ACC)	先行車軌跡追従	D121
			その他	D12X
		駐車支援	自動バレーパーキング	D131
			その他	D13X
		車線変更支援		D14
		合分流支援		D15
		右左折支援		D16
		自動発進/自動停止支援(信号機、停止線)		D17

表2　特許が所属する技術分析軸

文献(特許)	D121	D16	D17
JP-A-2011-141747		1	
JP-A-2011-141802	1		1

報告書の自動更新のためには，各特許を技術分析軸に自動分類することが必要不可欠である．そのため，本研究では，各特許を技術分析軸へ分類することを目的とし，特許の構造を明らかにする．

3. 関連研究

本節では，「技術文書の構造解析」と「日英対訳コーパス」に関する関連研究について述べる．

3.1 技術文書の構造解析

Prabhakaranら[1]は, 科学的なトピックの成長と衰退を予測するために，修辞的機能（方法，目標，結果など）とトピックの関係を調査した．WoSコーパスに含まれる，筆者が明示的にセクションラベルを用いた，自己ラベル付き抄録を利用して学習を行う. 1991年から2010年までのWeb of Scienceの240万件の抄録を対象に構文解析を行い, 方法や目的，結果など，7つのラベルを自動付与した．調査の結果，結果として議論されているトピックは衰退する傾向にあり，一方，方法論的な役割を果たしているトピックは成長の初期段階にあることが明らかになった．本研究では，パテントファミリーを利用して構造解析を行う．

Heffernanら[2]は, 論文に記載されたフレーズが問題や解決策を示しているかを判断する分類器を作成した．この研究ではまず，論文内から，”problem”と”solution”に類似した語句をそれぞれ収集し，その後，収集した語句を含む文の構文の解析を行い，問題や解決策を示しているといえるフレーズを正のサンプルとした．また，正のサンプルと同じ文型であり，なおかつ，問題や解決策を示していないフレーズを負のサンプルとした．収集したサンプルをもとに，Bag of Words などの様々な特徴量を組み合わせ，Naïve Bayes（NB），logistic regression（LR），support vector machine（SVM）の3種類の分類器を用いて分類を行った．訓練用データの作成にパテントファミリーを用いる点が本研究とは異なる．

Westonら[3]は，材料科学論文を対象に，事前に定義した7種類のタグを人手で論文の概要に付与し，そのデータを教師データとして，LSTMによる情報抽出を行った．本研究は，特許文献を対象に分析を行う点とタグ付与方法が異なる．

酒井ら[4]は，特許明細書における「発明の効果」から，技術課題情報を抽出する手法を提案した. 抽出方法として，文集合から「ができる」のような，技術課題情報を抽出するための手がかりとなる表現（以降，手がかり表現）を使用した．また，全ての手がかり表現を網羅したリストを人手で作成することは困難であることから，手がかり語に係る確率に基づくエントロピーを用いて自動的に収集した．本研究では手がかり表現を用いた抽出は行わず，機械学習により文の抽出を行う．

樽松[5]は，語句出現頻度を利用した特許公報からの課

題・手段推定システムを提案した．まず，専門家に課題と手段ごとに分類された特許公報に含まれる要約文から，課題について述べている文と手段について述べている文を抽出し，その後，抽出した文に出現する語句の出現頻度に基づいて文書のベクトル化を行った．そして，ベクトル化した文書の類似度に基づいて，課題・手段となる文を推定した．本研究は，米国特許を対象にしており，「効果」の抽出も行う．

文よりも短い単位で重要な情報を抽出する研究を紹介する．このような技術は，文書検索や，動向分析に役に立つとされる．本研究の目的である，特許出願技術動向調査報告書の自動作成においては，技術分析軸を抽出する必要があるため，将来的に非常に関わりの深い研究である．

Gupta ら[6]は，論文概要から「FOCUS」「TECHNIQUE」「DOMAIN」という 3 種類のカテゴリに属する語句をパターンマッチにより自動的に抽出する手法を提案した．この手法では，例えば「propose」の直後に出現する目的語は「FOCUS」に属する語句としている．

Nanba ら[7]は NTCIR-8 特許マイニングタスクにおいて，日本語と英語で書かれた特定分野の論文と特許内の，「効果」等に関する文を対象とし，要素技術と効果を自動抽出した．図 1 にタグ付与例を示す．図 1 のように，要素技術と効果を示す箇所にそれぞれ TECHNOLOGY タグと EFFECT タグを付与する．さらに, EFFECT タグの中には，属性を表す ATTRIBUTE と属性値を表す VALUE という 2 種類のタグが付与されている．そして，機械学習により，入力された文に対してこれらのタグを自動的に付与した．要素技術と効果を抽出することができれば，技術動向を把握することが容易になる．例えば，特許内の「効果」等に関する文を分類し, その後 Nanba らの技術を用いることで，より効率的に分析軸の抽出を行うことができる．

[Japanese]
PM 磁束制御用コイルを設けて<TECHNOLOGY>閉ループフィードバック制御</TECHNOLOGY>を施すため，<EFFECT><ATTRIBUTION>電力損失</ATTRIBUTION>を<VALUE>最小化</VALUE></EFFECT>できる．
[English]
Through <TECHNOLOGY> closed-loop feedback control </TECHNOLOGY>, the system could <EFFECT> <VALUE> minimize </VALUE> the <ATTRIBUTION> power loss </ATTRIBUTION> </EFFECT>.

図 1　要素技術と効果のタグ付与

福田ら[8]は，学術論文を対象とし，学術論文固有の特徴を用いた分類手法を提案した．まず，Nanba らの手法を用いて，論文データから要素技術・効果を抽出し，タグ付けをする．その後，タグ付けされた語句をそれぞれ抽出し，要素技術リスト，属性リスト，属性値リストを作成する．各リストを，k-NN 手法と SVM 手法において学習を行う際の素性とし，学術論文の分類を行った．その結果，各リストを用いない手法よりも高い精度の分類を行うことが可能なったと述べている．

谷中ら[9]は，課題が類似する，または解決策が類似する特許明細書をクラスタリングし，クラスタ内の特許文献に共通する技術課題を抽出する手法を提案した．まず,「本発明は…する」という表現を手がかり表現として，述語項構造解析ツールを用いて「本発明」を主語とする動詞と目的格を抽出し，技術課題の要点とする．そして，抽出された動詞と目的格で使用されている名詞の出現頻度(TF)と，全文献の技術課題の要点で使用されている名詞の出現頻度の対数(IDF) を掛け合わせた TF・IDF 値が大きい単語を，そのクラスタの技術課題としている．

3.2　日英対訳コーパス

本研究では，日本国特許内の文を英語に翻訳する手順がある．翻訳器を構築する際，一般的には対訳コーパスを用意する必要がある．本節では，コーパスの構築について述べる．

Nomoto ら[10]は，TALPCo と呼ばれる対訳コーパスを作成した．このコーパスでは，日本語に対して，ミャンマー語，マレー語，ビルマ語，インドネシア語，英語の翻訳を収録している．また，東京外国語大学のオープンランゲージリソースに収録されている基本語彙と例文をもとに構成されている．

今村ら[11]は，病院や商業施設，観光地等で活躍する翻訳システムを構築するための多言語対訳コーパス「GCP コーパス」を作成した．このコーパスは，医療，防災，ショッピング，観光の 4 分野にまたがり，アジア言語を含む 10 言語をカバーしている．

石坂ら[12]は，翻訳文書の著作権を考慮しつつ，Web で公開されているオープンソースソフトウェアのマニュアルを収集し，日英対訳コーパスを構築した．対訳文の総数は約 50 万文となっている．

上記で挙げたコーパスはいずれも人手で作成されており，膨大なコストがかかっている．翻訳器を構築する際，適切なコーパスを用いる必要があるが，自作することは困難である．本研究では，既存の日英対訳コーパスを用いて構築された翻訳器を用いる．使用する日英対訳コーパスに関する詳細は 4 節で述べる．

4.　米国特許の構造解析

4.1　パテントファミリーを用いた米国特許の構造解析

1 節で述べた通り，米国特許には，「効果」や「課題」などを記述する項目が明示的には設けられていないため，構

造解析が困難となっている．そこで，パテントファミリーを用いて構造タグ付き米国特許データセットを作成し，それを教師データとして機械学習させることで，米国特許の構造解析システムを構築する手法を提案する．

一般に，特許権は，国毎に独立して権利が付与される．各国で特許権を取得したい場合，特許出願人は，同じ発明を複数の国に特許出願する必要がある．このような，同一内容の特許文献群をパテントファミリーと呼ぶ．パテントファミリーは，言語や構造は異なるが，出願書類を構成する文章は，ほぼ対訳となっている．そこで，本研究では，「効果」や「課題」などの構造タグが付与された日本国特許内の文と意味の共通する米国特許内の文に同様の構造タグを付与することで，米国特許の構造を明らかにする．そして，構造化した米国特許を教師データとして機械学習を行い，米国特許の構造解析システムを構築する．

日本国特許の構造について述べる．日本国特許は，図 2 のようなフォーマットで構成されており，特許出願者はこのフォーマットに沿って出願書類を作成する．このような構造は，読者が得たい情報を効率的に獲得することを可能にする．例えば，[発明の効果]の欄を読むことで，その発明の「効果」を知ることができる．

一方で，米国特許には，このような構造は明示的には存在しない．そのため，米国特許の構造を解析する手法が求められている．

[書類名]
[発明の名称]
[技術分野]
[背景技術]
[先行技術文献]
[特許文献]
[非特許文献]
[発明の概要]
[発明が解決しようとする課題]
[課題を解決するための手段]
[発明の効果]
…

図 2　特許出願書類のフォーマット

4.2 米国特許の構造解析の手順

米国特許の構造解析手順を以下に示す.

（1）日本国特許において,「効果」,「課題」,「手段」の構造タグが付与された文を，それぞれ英語に翻訳.
（2）米国特許内の全ての文に対し，(1) の翻訳結果とのコサイン距離をそれぞれ算出. スコアが最も高くなる文に日本国特許と同じ構造タグを付与.
（3）（2）のデータを教師データとして機械学習を行い，米国特許における，「効果」,「課題」,「手段」に関する文を自動抽出.

（1）と（2）は 4.3 節で，（3）は 4.4 節で詳しく述べる．

4.3 データセット作成

まず,日英対訳コーパス JParaCrawl[13]と，シーケンスモデリングツール FAIRSEQ[14]を用いて構築された翻訳器で,日本国特許内の構造タグ付きの文を英語に翻訳する. なお, JParaCrawl とは，Web から文章をクロールすることにより収集した, 約 1000 万対の日英対訳を収録したコーパスである． また，FAIRSEQ は，機械翻訳などのテキスト生成モデルを訓練することができるツールであり，PyTorch で実装されている.

次に，翻訳した文と意味の共通する文を，米国特許から抽出する．まず，単語の出現頻度（TF）で文をベクトル化し，その後，日本国特許のパテントファミリーである米国特許内の全ての文に対して，翻訳結果の文とのコサイン距離を算出する. なお, ベクトル化の際, “a”や“the”といった, あまりに一般的である単語は，ストップワードとして排除する．そして，スコアが最も高い米国特許内の文を，日本国特許内の文と意味の共通する文であると見なし，日本国特許と同じ構造タグを付与する．この方法により，構造タグ付き米国特許データセットを作成する．パテントファミリーは忠実に翻訳されているので，スコアが最も高くなる文は，ほぼ対訳文となることが期待できる.

上記の手順により，日本国特許をパテントファミリーとして持つ米国特許 4,410 件に対して構造タグの付与を行った．図 3 に「効果」に関する日本国特許の 1 文と，それに対応する米国特許の 1 文の例を示す.

<JA>環状部材をワッシャとしてこれをケースとアイドラ歯車との間に装着することにより，リバースシャフトに段差を設けることが不要となる．</JA>
<EN>When the ring-shaped member used as a washer is mounted between the case and the idler gear, it becomes unnecessary to form any step portion on the reverse shaft.</EN>

図 3　「効果」に関する日本国特許の 1 文と
対応する米国特許の 1 文

人手でいくつかの対訳文を確認したところ，概ね正しい対訳が得られていた．一方で，表層的に類似していることにより，対訳として誤った文が最も高いスコアを算出し，正しい対訳文となっていないものもいくつか見られた.

4.4 機械学習による米国特許の構造解析

4.3 節で作成したデータセットを教師データとして機械学習を行う．分類器には，fastText[15]を使用する．fastText は，文をサブワードに分割することで，文をベクトル化する．サブワードとは，単語以下の単位である，文字や部分文字列のことであり，未知語が登場するテキストの分類にも有効であるとされている．

分類器を学習させることで，与えられた文が，「効果」，「課題」，「手段」に関する文かどうかを自動的に判断させる．機械学習を用いた構造タグの自動付与により，米国特許の構造を明らかにすることができる．

5. 実験

4 節で提案した手法の有効性を確認するため，実験を行った．

5.1 実験設定条件

実験データ

4.2 節により，「効果」，「課題」，「手段」の構造タグを自動付与した米国特許 713,658 文に対して，513,658 文を訓練用，200,000 文を評価用として用いた．付与された構造タグの内訳を表 3 に示す．

表 3　構造タグ付き文の内訳

	訓練用	評価用
効果	3,102	1,231
課題	9,459	3,686
手段	18,571	7,754
構造タグなし	482,526	187,329
計	513,658	200,000

パラメータ設定・評価方法

fastText のパラメータは，単語の次元数が 300 次元，epoch=30 で学習を行う．評価は構造タグそれぞれに対して，精度，再現率，F 値を算出する．

5.2 結果

構造タグごとの分類の評価を行った．表 4 に実験の結果を示す．実験の結果，全体で精度 0.0684，再現率 0.2819 が得られた．

表 4　構造タグごとの分類の精度

	精度	再現率	F 値
効果	0.0211	0.0024	0.0044
課題	0.0733	0.2235	0.1104
手段	0.0672	0.3540	0.1130
計	0.0684	0.2819	0.1101

5.3 考察

本研究では，単純なコサイン距離を用いて対訳文を決定したため，表層的に類似している文を対訳文としてしまうことがあった．「効果」，「課題」，「手段」を示す文に固有な表現に基づいて翻訳することができれば，精度の向上につながると考える．そのために，翻訳器構築の際に用いたコーパスのリソースを，本研究のような Web からではなく，特許文献をもとにした対訳コーパスを用いることが理想である．また，詳しい精度の調査と，トークナイズ手法の精査も必要である．

特許には，似ている文が 2 つ以上存在するようなこともあるが，本研究では，データセット作成の際に 1 つに特定してしまっている．翻訳文との類似スコアが高い英文であっても，類似スコアが最も高い文ではない場合は対訳文とされないため，本来は対訳文である文を取得できていない可能性がある．類似スコアによっては，対訳文を複数決定することを検討する必要がある．

実験の結果，「課題」，「手段」に比べて，「効果」の分類精度が低くなっている．この理由として，データセットにおける「効果」タグを付与された文の数が，「課題」，「手段」に比べて少ないからだと考えられる．分類の精度をさらに向上させるには，さまざまな機械学習手法を適用し，比較する必要がある．

6. おわりに

本研究では，米国特許 4,410 件に対し，構造タグを付与したデータセットを作成した．また，作成したデータセットを使用して，米国特許から，各特許を技術分析軸に分類するために重要となる文を自動分類する実験を行った．実験の結果，精度 0.0684，再現率 0.2819 で「効果」，「課題」，「手段」に関する文を分類できることがわかった．

7. 今後の課題

パテントファミリーにおける，意味の共通する文を対応させる精度の向上が必要である．現段階では，Web をリソースとしたコーパスを用いて機械翻訳を行っているため，今後は，特許専用コーパスを用いて翻訳モデルを作成する．また，翻訳精度の調査や，トークナイズ手法と類似度の算出方法の精査も併せて実施する．

米国特許の構造解析に関して，今後は，様々な機械学習手法を適用する．また，さらに，文の抽出結果を当初の予定である技術動向分析に応用する．その際には，文書単位ではなく，文単位での構造解析が必要となる．

謝辞　本研究は JSPS 科研費 JP19K12101 の助成を受けたものである。

参考文献

[1] Vinodkumar Prabhakaran, William L. Hamilton, Dan McFarland, and Dan Jurafsky. “Predicting the rise and fall of scientific topics from trends in their rhetorical framing,” In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 1170-1180 (2016)

[2] Kevin Heffernan and Simone Teufel. “Identifying problems and solutions in scientific text,” Scientometrics, 116(2), 1367-1382 (2018).

[3] Leigh Weston, Vahe Tshitoyan, John M. Dagdelen, Olga Kononova, Kristin Aslaug Persson, Gerbrand Ceder, and Anubhav Jain. “Named entity recognition and normalization applied to large-scale information extraction from the materials science literature,” Chemical Information and Modeling, 59(9), 3692-3702 (2019).

[4] 酒井 浩之, 野中 尋史, 増山 繁. “特許明細書からの技術課題情報の抽出” 人工知能学会, 24 巻・6 号, 531-540 (2009).

[5] 榑松 理樹. “語句出現頻度を利用した公開特許からの課題・手段推定システムの検討” 2016 年度人工知能学会全国大会 (2016)

[6] Sonal Gupta and Christopher D. Manning. “Analyzing the dynamics of research by extracting key aspects of scientific papers,” In Proceedings of 5th International Joint Conference on Natural Language Processing, 1-9 (2011).

[7] Hidetsugu Nanba, Atsushi Fujii, Makoto Iwayama, and Taiichi Hashimoto. “Overview of the patent mining task at the NTCIR-8 workshop,” In Proceedings of NTCIR-8 Workshop Meeting, 293-302 (2010).

[8] 福田 悟志, 難波 英嗣, 竹澤 寿幸. “要素技術とその効果を用いた学術論文の自動分類” 日本図書館情報学会誌, 62 巻・3 号, 145-162 (2016).

[9] 谷中 瞳, 大澤 幸生. “特許文献を利用した技術課題の抽象化方法の検討” 2016 年度人工知能学会全国大会 (2016).

[10] Hiroki Nomoto, Kenji Okano, David Moeljadi, and Hideo Sawada. “TUFS Asian language parallel corpus,” 言語処理学会 第 24 回年次大会, 436-439 (2018).

[11] 今村 賢治, 隅田 英一郎. “グローバルコミュニケーション計画のための多言語パラレルコーパス” 言語処理学会 第 24 回年次大会 発表論文集, 512-515 (2018).

[12] 石坂 達也, 内山 将夫, 隅田 英一郎, 山本 和英. “大規模オープンソース日英対訳コーパスの構築” 研究報告音声言語情報処理, 1-6 (2009).

[13] 森下 睦, 鈴木 潤, 永田 昌明. “JParaCrawl: 大規模 Web ベース日英対訳コーパス” 言語処理学会 第 26 回年次大会, 461-464 (2020).

[14] Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. “FAIRSEQ: A fast, extensible toolkit for sequence modeling,” In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics, 48-53 (2019).

[15] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. “Enriching word vectors with subword information,” In Proceedings of the 2017 Transactions of the Association for Computational Linguistics, 135-146 (2017).