

## ストーリー展開と一貫性を同時に考慮した歌詞生成モデル

## Modeling Theme and Story Line for Lyrics

渡邊 研斗 \*1  
Kento Watanabe松林 優一郎 \*1  
Yuichiro Matsubayashi乾 健太郎 \*1  
Kentaro Inui深山 覚 \*2  
Satoru Fukayama中野 倫靖 \*2  
Tomoyasu Nakano後藤 真孝 \*2  
Masataka Goto\*1 東北大学  
Tohoku University\*2 産業技術総合研究所  
AIST

By adopting recent advances in music creation technologies, people can now create songs in their personal computers. Computers can also assist in creating lyrics or generating them automatically, although this aspect has been less thoroughly researched and is limited to rhyme and meter. This study focuses on the structural relations in Japanese lyrics. We present novel generative models that capture the theme such as "Love", and capture the topic transitions between units peculiar to lyrics, such as *verse A*, *verse B* and *chorus*. This structure is modeled by combining Mixture of Unigram Model and Hidden Markov Model for representing topic transitions. To evaluate that our method models topic transitions, we evaluate the models by block ordering test. The result shows that the proposed models successfully capture the topic ordering. In the result of model analysis, we found that our approach captures the typical popular music story line.

## 1. はじめに

ポピュラー音楽において、歌詞は感情やメッセージ性を伝えるための重要な要素である。特にプロの作詞家は聴き手にメッセージを伝えやすくするために、(1)「愛」や「夏」のような一貫したテーマ、(2)「A メロ-B メロ-サビ」のようなブロック構造に表れる「話の場面設定-回想-主張、感情」のようなストーリー展開を考慮することが重要であると指摘している [1]。そのため、作詞過程におけるテーマやストーリー展開をモデル化することができれば、作詞支援技術や歌詞の構造解析技術へ応用できると考えられる。

しかし、既存研究の多くは韻やアクセント、ムードに着目しており、歌詞のテーマやストーリー展開を数理的にモデル化した研究少ない [2, 3]。また、経験的に指摘されていた「作詞におけるテーマ・ストーリー展開の重要性」を統計的に検証した研究は少ない [4]。

本研究では歌詞の内容を重要視した作詞支援システム実現への第一歩として、一貫したテーマとストーリー展開を考慮したモデルを提案する。実験では、歌詞内のブロックを正しい順序に並べるタスクを行い、提案モデルが楽曲のストーリー展開をモデル化する上で優れていることを示した。さらに「作詞におけるテーマ・ストーリー展開の重要性」を確認するために、学習されたモデルを定性的に分析した。

## 2. 提案手法

プロの作詞家は作詞における「一貫したテーマ」や「場面の展開」の重要性を主張している [1]。そのため、歌詞に含まれる感情やメッセージを数理的にモデル化するためには、(a) 歌詞の一貫したテーマと、(b) 歌詞のストーリー展開をモデル化することが自然だと考えられる。本研究では (a) に対して Nigam らの混合ユニグラムモデル (Mixture of Unigram

Model (MUM)) [5] を利用し、(b) に対して Barzilay らのコンテンツモデル (Content Model (CM)) [6] を応用した。

MUM では 1 つの歌詞が 1 つのテーマ  $y$  を持つ。 $y$  はテーマ分布  $\theta_{theme}$  により生成され、歌詞内全ての単語  $w$  は生成確率分布  $\phi_y$  から生成される (図 1(a))。このモデルにより、一貫したテーマをモデル化する。CM では、1 つのブロック  $b$  が 1 つの隠れ状態  $s$  を持ち、各ブロック  $b$  の隠れ状態  $s_b$  は遷移分布  $\theta_{s_{b-1} \rightarrow}$  により、直前ブロック  $s_{b-1}$  から遷移する。ここでブロックとはポピュラー楽曲における「A メロ」や「サビ」などの楽曲構造を表し、本研究では歌詞カードに書かれている段落をブロックと仮定する。また、隠れ状態  $s$  をストーリー要素と呼び、 $s$  の遷移列  $s_1, s_2, \dots$  をストーリー展開と呼ぶ。各ブロック  $b$  の単語  $w$  は生成確率分布  $\phi_{s_b}$  から生成される (図 1(b))。このモデルによりブロックごとのストーリー要素の遷移構造をモデル化する。

ただし、MUM は曲全体の一貫性を考慮しているが、ストーリーの順序関係を考慮していない。また、CM におけるストーリー展開は 1 次マルコフ連鎖のため、ブロックの長さが大きい場合、曲全体の一貫性が考慮できない。そこで本研究では、MUM と CM を組み合わせ、全体の一貫性を考慮したストーリー展開をモデル化する。

## 2.1 MUM-CM: 混合ユニグラム-コンテンツモデル

MUM-CM のプレートノテーションと生成過程を図 1(c) と図 2 にそれぞれ示した。ここで  $I$  と  $J$  はそれぞれテーマ数とストーリー要素数を表し、 $m$  は歌詞のインデックス、 $b$  は各歌詞のブロックの位置を表す。 $\theta, \phi, \lambda$  は確率分布を表し、 $y, s, x$  は潜在変数を表す。また  $\alpha, \beta, \eta, \zeta, \epsilon$  は各確率分布の平滑化パラメータである。 $w$  は観測値である単語を表す。

このモデルでは、MUM と CM を独立したモデルとして扱い、二値の潜在変数  $x$  の値によって使用するモデルを選択する。この  $x$  は歌詞内の各単語に割り当てられ、ベルヌーイ分布  $\lambda_m$  によって 0 か 1 の値を持ち、 $x$  が 0 のときは MUM から単語を生成し、1 のときは CM から単語を生成する。以下

連絡先: 渡邊研斗, 東北大学, 宮城県仙台市青葉区荒巻字青葉  
6-6-05, 022-795-7091, kento.w@ecei.tohoku.ac.jp

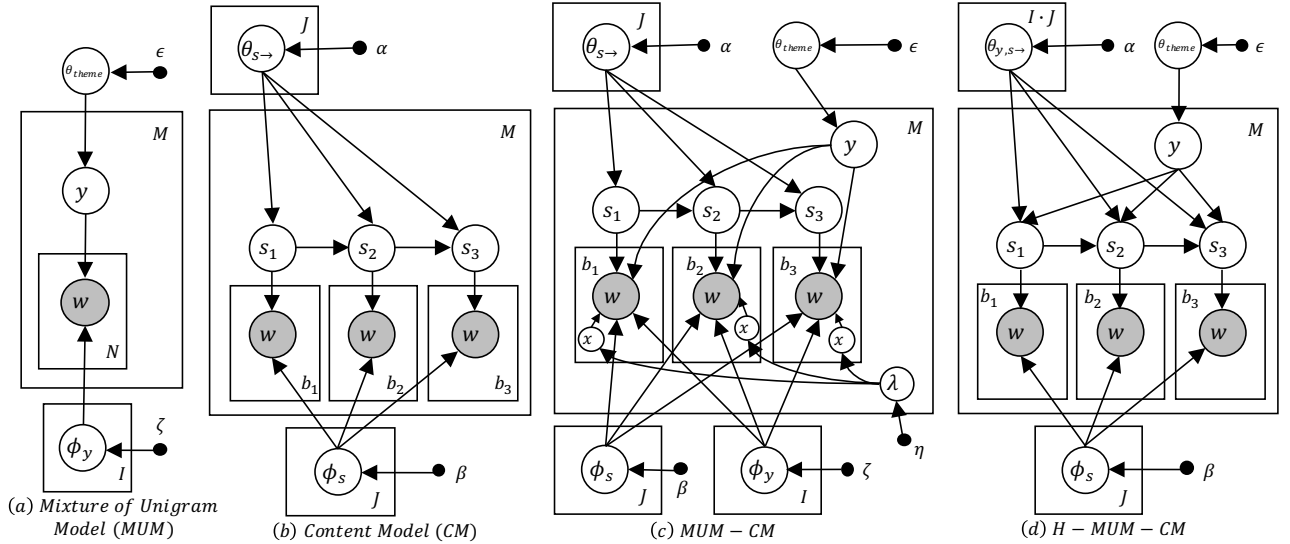


図 1: 既存手法 (a)(b) と提案手法 (c)(d) のプレートノーテーション

Draw  $\theta_{theme} \sim Dir(\epsilon)$   
 For each theme  $y = 1, 2, \dots, I$ : draw  $\phi_y \sim Dir(\zeta)$   
 For each story  $s = 1, 2, \dots, J$ :  
   Draw  $\theta_{s \rightarrow} \sim Dir(\alpha)$   
   Draw  $\phi_s \sim Dir(\beta)$   
 For each lyrics  $m = 1, 2, \dots, M$ :  
   Draw  $y_m \sim Multi(\theta_{theme})$   
   Draw  $\lambda_m \sim Beta(\eta)$   
   For each block  $b = 1, 2, \dots, B_m$ :  
     Draw  $s_b \sim Multi(\theta_{y_m, s_{b-1} \rightarrow})$   
     For each word  $w$  in block  $b$ :  
       Draw  $x \sim Bernoulli(\lambda_m)$   
       Draw if  $x = 0$ : draw  $w \sim Multi(\phi_{y_m})$   
       Draw if  $x = 1$ : draw  $w \sim Multi(\phi_{s_b})$

図 2: MUM-CM の生成過程

に MUM-CM における歌詞の生成確率を示す。

$$P(lyrics_m) = P(x=0|\lambda_m) \sum_{y=1}^I P(y|\theta_{theme}) \prod_{w \in m} P(w|\phi_y) \\ + P(x=1|\lambda_m) \sum_{s_{all}} \prod_{b=1}^{B_m} P(s_b|\theta_{y, s_{b-1} \rightarrow}) \prod_{w \in b} P(w|\phi_{s_b}) \quad (1)$$

ここで、 $B_m$  は歌詞  $m$  内のブロック数である。 $s_{all}$  は起こりうる全てのストーリー展開である。 $b=1$  のとき、 $\theta_{s_0 \rightarrow}$  は初期状態確率を表す。このモデルにより、曲全体の一貫性が重要な部分では MUM を使用し、ストーリー展開が重要な歌詞では CM を使用することが期待される。

## 2.2 H-MUM-CM: 階層的混合ユニグラム-コンテンツモデル

前節の MUM-CM ではテーマ  $y$  とストーリー展開  $s_1, \dots, s_B$  が独立に生成されるモデルであった。しかし、例えばテーマが「愛」なら恋愛らしいストーリー展開になるはずである。本研究ではテーマごとにストーリー展開を生成するモデルとし

Draw  $\theta_{theme} \sim Dir(\epsilon)$   
 For each story  $s = 1, 2, \dots, J$ :  
   Draw  $\phi_s \sim Dir(\beta)$   
   For each theme  $y = 1, 2, \dots, I$ :  
     Draw  $\theta_{y, s \rightarrow} \sim Dir(\alpha)$   
 For each lyrics  $m = 1, 2, \dots, M$ :  
   Draw  $y_m \sim Multi(\theta_{theme})$   
   For each block  $b = 1, 2, \dots, B_m$ :  
     Draw  $s_b \sim Multi(\theta_{y_m, s_{b-1} \rightarrow})$   
     For each word  $w$  in block  $b$ :  
       Draw  $w \sim Multi(\phi_{s_b})$

図 3: H-MUM-CM の生成過程

て階層的混合ユニグラム-コンテンツモデル (H-MUM-CM) を構築した。H-MUM-CM のプレートノーテーションと生成過程を図 1(d) と図 3 に示した。本モデルは最初にテーマ  $y$  を生成し、直前のストーリー要素  $s_{b-1}$  と各  $y$  ごとに用意されたストーリー遷移分布  $\theta_{y, s_{b-1} \rightarrow}$  から次のブロックのストーリー要素  $s_b$  を生成する。以下に H-MUM-CM における歌詞の生成確率を示す。

$$P(lyrics_m) = \sum_{y=1}^I P(y|\theta_{theme}) \\ \sum_{s_{all}} \prod_{b=1}^{B_m} P(s_b|\theta_{y, s_{b-1} \rightarrow}) \prod_{w \in b} P(w|\phi_{s_b}) \quad (2)$$

このモデルにより、曲全体のテーマごとにストーリー展開を考慮できることが期待される。

## 3. 並び替え問題による評価

本節では、提案モデルがブロックのストーリー展開を正しく学習しているか評価するために、並び替え問題により評価を

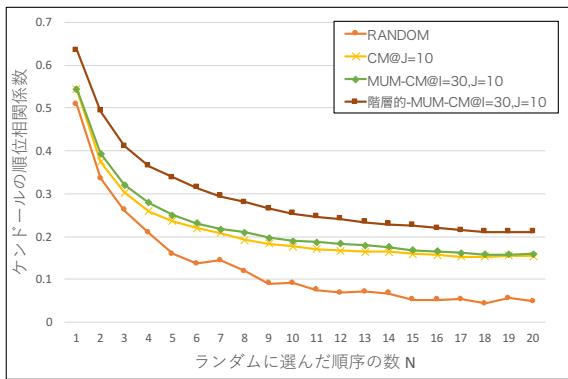


図 4: 各モデルごとのケンドールの順位相関係数の平均

行った。もし提案手法が歌詞内のブロックを適切に並べ替えることができれば、歌詞全体のストーリー展開をモデル化できたことになる。評価の方法として、全ての可能な順序に対してモデルの生成確率が最も高いものを選んだ場合を考え、正解の並び順との順位相関を計る。ただしこの場合、最大で  $B!$  通りの順序を探索する必要があるため、計算量の観点から評価が難しい。そこで、Zhai らの手法 [7] で近似することで、この問題に対処した。以下に並び替え問題の具体的な手順を示す。

1. テストデータの各歌詞  $m$  について、ブロックの順列からランダムに  $N$  個の順序を選ぶ。
2. 選んだ各順序について、生成確率  $P(\text{lyrics})$  を計算する。
3. 最も生成確率が高い順序をモデルが選択した順序  $A'$  とする。
4. モデルが選択した順序  $A'$  と正しい歌詞の順序  $A$  を比較し、ケンドールの順位相関係数  $\tau(3)$  を計算する。

$$\tau = \frac{c^+(A, A') - c^-(A, A')}{T(T-1)/2} \quad (3)$$

ここで  $c^+(A, A')$  は  $m$  中の任意の二つのブロックについて、 $A$  における順序と  $A'$  における順序が一致した回数であり、 $c^-(A, A')$  は一致しない回数である。ケンドールの順位相関係数  $\tau$  は  $-1$  から  $1$  までの値を取り、 $1$  に近づくほど 2 つの順序が相関を持ち、 $-1$  に近づくほど逆の相関を持つことを表す。また、本研究では  $N = 20$  と設定した。

### 3.1 訓練・評価データ、パラメータチューニング

本研究では、予めブロックに分割されている日本のポピュラー音楽 16,563 曲の歌詞を用意した。このデータを訓練データ、パラメータ調整用の開発データ、テストデータの割合が 6:2:2 となるよう分割して用いた。

各モデルの推定にはギブスサンプリングを用い、サンプリングの回数を 1000 回とした。実験では、品詞が「名詞」「動詞」「形容詞」の内容語のみを使ってストーリー展開を学習する。このために、日本語の形態素解析器として MeCab[8] を、英語の形態素解析器として GENIA Tagger[9] を使用した。ハイパーパラメータ  $\alpha, \beta, \eta, \zeta, \epsilon$  は、コーパス全体の単語集合  $W$  と潜在変数  $y, s, x$  との同時確率 (例えば  $P(W, y|\alpha)$ ) を最大化するように EM アルゴリズムを用いて推定した。テーマ数  $I$  とストーリー要素数  $J$  を決定するために、 $I$  と  $J$  を 10 ずつ動かして、開発データを使った並び替え問題を行った。その結果、比較対象の全てのモデルに関して  $J = 30, I = 10$  が最も性能が良かったため、これらのパラメータを採用した。

### 3.2 結果

我々は並び替え問題における比較対象として、全ての可能なブロックの順序の中からランダムに順序  $A'$  を選ぶ「弱いベースライン」と、CM を用いた「強いベースライン」を用意した。図 4 に、各手法ごとのケンドールの順位相関係数の相対平均を示す。この表より、提案手法である H-MUM-CM が相対的に高い順位相関係数であり、テーマごとにストーリー展開をモデル化した方がブロックの順序性をモデル化できていることがわかる。また、Tukey-Kramer 検定による多重比較検定を行った結果、 $H=MUM=CM$  と他のモデルに有意な差があった。一方 MUM-CM と CM に有意な差がなく、同等の性能であった。これは、MUM-CM は順序を並び替えることのできない MUM と順序を並び替えることができる CM を独立したモデルとして組み合わせしており、CM と同等の並び替え性能しか持たないためだと考えられる。

## 4. 定性的評価

実際の作詞ではテーマやストーリー展開を考慮する必要があると経験的に言われているが、この経験則を統計的に確かめた研究はない。本節では、並び替え問題で高性能であった H-MUM-CM ( $I=30, K=10$ ) の単語生成確率とテーマごとのストーリー遷移確率を分析することで、この経験則が実際に成り立っているかを確認した。

### 4.1 ストーリー要素ごとの単語生成確率の分析

提案モデルがストーリー展開を適切にモデル化しているかを確認するために、遷移確率を分析したい。その前段階として本節では、遷移確率の分析を進めやすくするために、表 1 にストーリー要素  $s$  と各単語生成確率が大きい単語集合、我々が付与したストーリー要素のラベルを示した。

ストーリー要素  $s = 2$  は「行く」や「食べる」などの日常的な動作や「仕事」や「仲間」などの単語があることから「親近感・生活」というラベルを付与した。 $s = 3$  は「風」「街」「雨」などの現象や物体を表す名詞と「遠い」「咲く」などの物体の特徴や動作を表すや形容詞・動詞で形成されているため、「情景」というラベルを付与した。 $s = 5$  は英単語で形成されているが「love」や「lonely」などの恋愛に感傷有する単語のため「恋愛・英語」というラベルを付与した。その他のストーリー要素も単語の特徴から筆者らがラベルを付与した。以降、これら付与ラベルを使用してストーリー展開の分析を進める。

### 4.2 テーマごとのストーリー遷移確率の分析

本研究で提案した H-MUM-CM は 30 個のテーマごとに別々のストーリー展開をモデル化した。本節ではその中から、特にわかりやすいストーリー展開を学習した 3 つのテーマについて分析した。図 5 は各テーマごとの起こりやすいストーリー展開をグラフで表現したものである。ここで各ノードは表 1 のストーリー要素を表しており、「START」は曲の初期状態を表す潜在変数  $s_0$  である。また、矢印はストーリー要素の遷移を表しており、各数値は遷移確率  $P(s'|\theta_{y,s \rightarrow})$  を表す。

図 5 を見ると、テーマ  $y = 5$  では「情景」から「恋愛・切なさ」に遷移しやすく、ポピュラーな恋愛楽曲で見られるストーリー展開が学習されていることがわかる。また、テーマ  $y = 6$  では、「情景」から「恋愛・交際」を経由して「英語」へ遷移しやすく、英語のサビを含んだ恋愛楽曲のストーリー展開が学習された。テーマ  $y = 9$  では、「恋愛・交際」と「親近感・生活」を相互に遷移しやすく、Hip/Hop や Rap などで見られやすいストーリー展開が学習された。このように、ストーリー展

表 1: ストーリー要素  $s$  と単語生成確率  $P(w|\phi_s)$  上位 15 単語

s	付与ラベル	単語生成確率 $P(w \phi_s)$ 上位 15 単語
2	親近感・生活	行く, やる, 今日, 気, いう, 仕事, 無い, 人生, 出る, 持つ, 頭, 仲間, 日本, やめる, 食べる
3	情景	空, 風, 街, 雨, 花, 遠い, 消える, 星, 海, 月, 流れる, 白い, 雪, 咲く, 吹く
5	恋愛・英語	love, night, kiss, good, girl, dance, rain, hold, sweet, lonely, no, bye, stay, smile, moon
8	恋愛・切なさ	愛, 今, 心, 胸, 涙, 言葉, 忘れる, 抱きしめる, 愛す, 信じる, 想い, 強い, 優しい, すべて, 変わる
9	英語	get, know, say, baby, time, take, heart, feel, way, life, party, come, make, need, see
10	恋愛・交際	言う, いい, 好き, 人, 見る, 笑う, 泣く, 知る, 思う, 女, わかる, 顔, 男, 来る, 会う

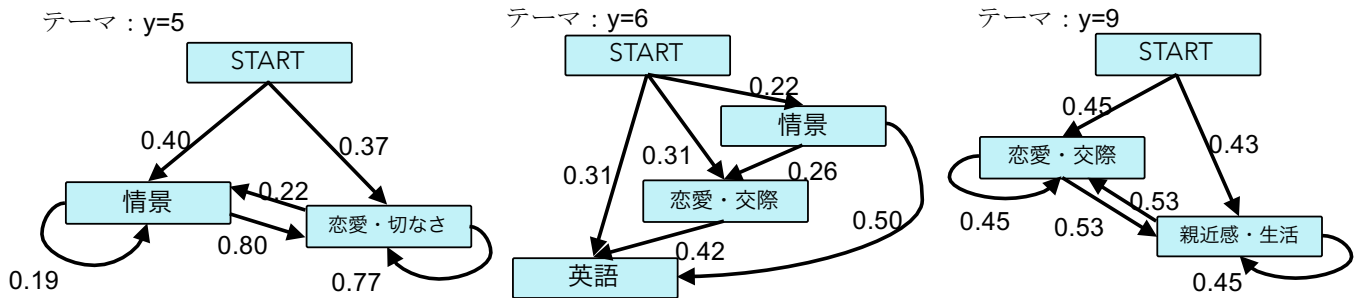


図 5: テーマごとの起こりやすいストーリー展開

開が異なるテーマごとに学習されていることがわかった。すなわち、今まで経験的に指摘されていたテーマやストーリー展開の重要性が統計的に示された結果となった。

## 5. おわりに

本研究では歌詞において重要とされている「一貫したテーマ」と「ストーリー展開」を考慮した歌詞生成モデルを実装した。具体的には一貫したテーマを混合ユニグラムモデルで、ストーリー展開をコンテンツモデルで表現し、この2モデルを組み合わせることで2つの要素を同時に扱うモデルを実現した。

提案手法がストーリー展開をモデル化できたかを確認するために並び替え問題を行った結果、テーマごとにストーリー展開を生成する提案手法が有意に高い性能を示し、ストーリー要素の順序性を学習できていることが確認できた。

各ストーリー要素の生成確率が大きい単語を分析した結果、<情景> <恋愛・切なさ>などのストーリー要素が学習されたことが確認できた。更に、各テーマごとのストーリー展開を分析した結果、<情景> <恋愛・切なさ>のようなポピュラーな恋愛楽曲に見られるストーリー展開が学習されるなど、テーマごとにストーリー展開が異なることを示した。この事実は作詞における「一貫したテーマやストーリー展開を考慮することは重要である」という経験則を確かめたことを意味する。

今後は、提案モデルの応用研究として、テーマやストーリー展開を入力するだけで膨大な可能性の中から単語列を探索する作詞システムを実現していく予定である。

## 参考文献

- [1] いしわり淳治, 吉元由美, みろく, zopp, MiChi, ヒロイズム, 宇多丸, 森浩美. 作詞のための8の極意. yamaha music media corporation, 2014.
- [2] Rada Mihalcea and Carlo Strapparava. Lyrics, music, and emotions. In *Proc. of MNLN-CoNLL 2012*, pp. 590–599, 2012.
- [3] Karteek Addanki and Dekai Wu. Unsupervised rhyme scheme identification in hip hop lyrics using hidden markov models. In *Statistical Language and Speech Processing*, Vol. 7978 of *Lecture Notes in Computer Science*, pp. 39–50, 2013.
- [4] Kento Watanabe, Yuichiroh Matsubayashi, Kentaro Inui, and Masataka Goto. Modeling structural topic transitions for automatic lyrics generation. In *Proc. of PACLIC 2014*, pp. 422–431, 2014.
- [5] Kamal Nigam, Andrew Kachites McCallum, Sebastian Thrun, and Tom Mitchell. Text classification from labeled and unlabeled documents using em. *Machine learning*, Vol. 39, No. 2-3, pp. 103–134, 2000.
- [6] Regina Barzilay and Lillian Lee. Catching the drift: Probabilistic content models, with applications to generation and summarization. In *Proc. of NAACL 2004*, pp. 113–120, 2004.
- [7] Ke Zhai and Jason D. Williams. Discovering latent structure in task-oriented dialogues. In *Proc. of ACL 2014*, pp. 36–46, 2014.
- [8] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *Proc. of EMNLP 2004*, pp. 230–237, 2004.
- [9] Yoshimasa Tsuruoka, Yuka Tateishi, Jin-Dong Kim, Tomoko Ohta, John McNaught, Sophia Ananiadou, and Junichi Tsujii. Developing a robust part-of-speech tagger for biomedical text. *Advances in Informatics - 10th Panhellenic Conference on Informatics*, pp. 382–392, 2005.