単語アライメントの誤り対応を用いた歌ことばのコノテーション検出

陳 旭東・山元 啓史（東京工業大学 環境・社会理工学院）
ホドシチェク ボル（大阪大学 大学院人文学研究科）

概要：本稿では和歌とその現代語訳の単語アライメントの誤った対応を用いた古今和歌集における歌ことばのコノテーション抽出の手法について報告する．提案手法は，対象単語について，文単位で現代語訳で補足されたノンリテラル要素（和歌の原文には出現せず現代語訳にのみ出現する意味の単位）を単語アライメントで誤って（不適切に）対応づけられた要素で定義することで，コノテーションの候補を検出する．検出された候補は，具体的にはどの和歌でどの訳者によるものかを確認でき，総合的にはネットワーク図に可視化し俯瞰できるような仕様にしている．植物類の歌ことば群に対し，実験を行った結果，単語アライメントの誤り対応で抽出されるノンリテラル要素は，現代語文を成立させる過程で生じた機能的な用語と，コノテーションに相当すると思われる用語であった．これらの要素を可視化することにより，訳者間に共通した要素と訳者によって異なる要素が描画できた．
キーワード：歌ことば，翻訳，単語アライメント，コノテーション

Connotation Detection for Classical Poetic Japanese Vocabulary Using Word Alignment Mismatch

Xudong Chen / Hilofumi Yamamoto (School of Environment and Society, Tokyo Institute of Technology)
Bor Hodošček (Graduate School of Humanities, Osaka University)

Abstract: This paper proposes a method for extracting the connotation of classical poetic Japanese vocabulary in the *Kokinshū* from words misaligned between the classical Japanese poems and their ten contemporary translations. The method defines connotation candidates of a poetic word as non-literal elements (misaligned words in a word alignment model) in the translations of each poem where the poetic word occurred. The method is characterized with its transparency (we can check from which poem and by which translator the connotation candidates occur) and distant readability (we can visualize the candidates in a network for a bird's-eye view). An experiment on classical poetic flora words showed that the non-literal elements extracted from misaligned words were both functional words necessary for the realization of the poem in contemporary Japanese and the aforementioned connotative words. In the network visualization, hubs were commonly supplemented non-literal elements among translators, that is, commonly acknowledged connotations; periphery vertices were non-literal elements that varied among translators.
Keywords: classical poetic Japanese, translation, word alignment, connotation

1. はじめに

本稿では，和歌と現代語訳の単語アライメントの誤った, または不適切な対応（以下「誤り対応」[*1]）を用いた歌ことばのコノテーション抽出の手法について報告する．単語アライメントはパラレルテキストにおける対訳語を自動的に対応づける言語処理の技術である．誤り対応は，対訳語を除いて単語アライメントモデルに対応づけられる統計的に関連性の高い非対訳語とする．

和歌の原文に出現しないのに，各現代語訳において共通して追加される要素（原文にとってはノンリテラルな要素）がある．本稿では，このようなノンリテラルな要素を Schramm のモデル [13] で捉え直し，特定の歌ことばについて単語アライメントモデルの誤り対応を意図的に用いて抽出する．抽出した要素が対象となる歌ことばのコノテーションとして捉えられるかを考察する．

植物類の歌ことば 6 語に対し実験を行った結果，単語アライメントの誤り対応で抽出されるノンリテラル要素は，現代語文を成立させる過程で生じた機能的な用語（指示語，代名詞，1 語の出現に対して複数回訳された語）と，コノテーションに相当する用語（後述）であった．これらの要素をネットワーク図で可視化した結果，訳者間に共通したものと訳者によって異なるものが描画できた．

[*1] いわゆる「正しい対応」の余事象である．

2. 関連研究

2.1 Schramm のコミュニケーション・モデルに基づくコノテーションの捉え方

Schramm のモデル [13] では，コミュニケーションは

送信元，エンコーダー，信号，デコーダー，受信者先の 5 つの要素から構成されている．送信元がメッセージを信号にエンコードし，その信号を受信者がデコードする．送信元と受信者はそれぞれ独自の経験野（field of experience）を持ち，両者の経験野の重なりが大きければ大きいほど受信者が信号を正しくデコードできる．

文献 [15] はコノテーションの定義を Schramm のコミュニケーション・モデルをヒントにして，和歌研究者による現代語訳から歌ことばのコノテーション（文字として明示されない要素[*2]）を抽出した．この研究は，現代人が和歌を読むコミュニケーション・モデルを (a) 歌人が和歌を書いて専門家がそれを読むサブ・モデルと (b) 専門家が和歌を現代語訳にして一般読者がその訳を読むサブ・モデルに分けた（図 1）．一般読者と歌人の間では経験野の重なりが少なく，歌人と専門家の間 (a) そして専門家と一般読者の間 (b) では経験野の重なりが大きい．したがって，専門家がサブ・モデル (a) で明示的に言語化されない・言語化する必要のない信号（和歌の原文）を，サブ・モデル (b) で一般読者がより正しく理解できるような信号（現代語訳）に加工しなければならない．そこで，(a) と (b) の信号の差分で (a) の経験野では言語化されなかった内容（コノテーション）が得られる．

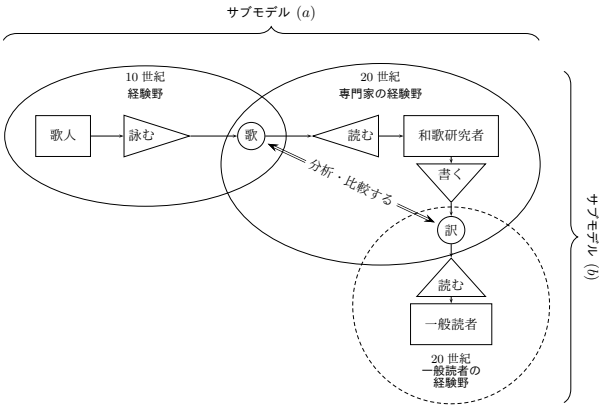


図 1 Schramm のモデルで捉える歌と現代語訳：信号として歌と現代語訳の比較でコノテーション相当の要素を示すことができる [15].

Schramm のモデルに基づくコノテーションの捉え方のメリットは，一般的には概念的にしか定義できないデノテーションとコノテーションの境界線を訳語[*3]の集合として定義し，コノテーションとして可視化できることである．

[*2] 言外の意とも言われる．
[*3] 現代語訳から古語相当の要素を引いた残差に相当する訳語．

2.2 差分モデル

コノテーションを 2 つのコミュニケーション・モデルにある信号の差分で捉える観点から，和歌の現代語訳から原文の語を引き算する手法が 2 つ提案されている [16], [17].

形式的に，[17] は，各和歌 S_i [*4]とその現代語訳 T_i[*5]に対し，その単語リストの差分 ($R_i = T_i \setminus S_i$) を求めることで，現代語訳文におけるノンリテラルな要素の抽出する方法で，[16] は和歌と現代語訳の言語空間における語の閾値以上の共出現パターンを和歌のグラフ G_s[*6]と現代語訳のグラフ G_t [*7]で描画し，両者のエッジの差分[*8]を求めグラフを再構築する方法である．

ただし，[16] では，和歌ごとの現代語訳で追加されるノンリテラル要素の集合 (R_i) を具体的に示せるが，R_i にある要素が原文 S_i にあるどの歌ことばのコノテーションなのかが明確ではなかった．一方で，[17] は，共出現パターンのリストの差分を求めることで，ノンリテラル要素が由来する文脈の推測ができるようにしているものの，和歌ごとにすこし異なる歌ことばのコノテーションのバリエーションまでは据えられなかった[*9]．また，2 つの手法では，本来同一視できない和歌の語彙と翻訳の語彙を分類語彙表のメタコードにより統一処理された．

そこで，本稿では，対象となる歌ことばについて，それが出現した対訳文のそれぞれにおいて当該歌ことばのアライメントの誤り対応という限定をかけながらノンリテラル要素を求め，コーパス全体で個別に求めた差分をまとめて俯瞰する手法を提案する．これによって，個別の和歌のみに出現する当該歌ことばのノンリテラルな要素も含め，抽出できる．[17] に比べ，この方法の利点は，特定の和歌–訳文対においてノンリテラル要素の対応先の歌ことばを推測できることである．なお，和歌にある語と翻訳にある語の意味の同定については，先行研究と異なり，分類語彙表のような知識基盤のコーディングに依存しない．

[*4] $S_i = \{s_n | n \in \mathbb{Z}\}$；$i$ は和歌の歌番号を指し，s_n は歌 S_i の n 番目の語を指す．
[*5] $T_i = \{t_m | m \in \mathbb{Z}\}$；前掲同様，$i$ は和歌の歌番号を指し，t_n は歌 S_i の翻訳 T_i の n 番目の語を指す．
[*6] $G_s = (V_s, E_s)$；V_s は対象となる歌ことばを含む和歌にある語に相当する頂点の集合である；E_s は対象となる歌ことばを含む和歌にある語の関わり合いに相当するエッジである．
[*7] $G_t = (V_t, E_t)$；前掲と同様に，V_t は対象となる歌ことばを含む和歌の翻訳にある語に相当する頂点の集合である；E_t は対象となる歌ことばを含む和歌の翻訳にある語の関わり合いに相当するエッジである．
[*8] つまり $E_t \setminus E_s$
[*9] すなわち E_r はつねに対象となる語と直接・間接的に関わるものの，E_r にある要素はどの歌のどの訳者から抽出されたかは明瞭に見えない．

3. 方法

3.1 材料

材料と材料の前処理について説明する．

古今和歌集と 1933 年から 1988 年にかけて公刊された現代語訳 10 種（表 1）から構築したパラレルコーパスを用いる．古今和歌集のデータは，[2] のデータを用いる．現代語訳のデータは，形態素解析済みで，[2] と同一のフォーマットにしたものを用いる．前処理として記号類をすべて除外する．また，2 資料における語が複合語である場合，その decomposition も格納されている．それら decomposition を省く．前処理したデータ量の詳細は，表 1 に示す．

表 1 データの詳細

	token 数	type 数	文書数
金子 [3]	42439	3356	1000
片桐 [4]	36362	2882	1000
小島・新井 [5]	33867	2955	1000
小町谷 [6]	30869	2692	1000
窪田 [7]	32210	2701	1000
久曽神 [8]	34050	2770	1000
松田 [9]	31860	3007	1000
奥村 [11]	32321	3153	1000
小沢 [12]	36173	3384	1000
竹岡 [14]	29844	2861	1000
total	339995	8252	10000
古今集データセット [2]	16687×10	1496	1000×10

3.2 Step 1：単語アライメントモデルの学習

単語アライメントモデルの IBM Model 2[1] を和歌の原文 → 現代語訳文の順方向と，現代語訳文 → 原文の逆方向で 2 つ学習させる．以下，IBM Model とその学習の設定について説明する．

IBM のモデルは，ソース言語のテキスト f からターゲット言語のテキスト e への翻訳タスクをノイズチャネルモデルとして組む．f から e への翻訳確率，すなわち f が与えられた時の e の確率は，ベイズの定理に従って式 (1) で計算する．

$$p(e|f) = \frac{p(e)p(f|e)}{p(f)} \tag{1}$$

$p(\cdot)$ は言語モデルを指す．そして f にとってもっとも適切な翻訳 $\hat{e}$ は式 (2) で計算する:

$$\hat{e} = \arg\min_{e} p(e)p(f|e) \tag{2}$$

$p(e)$ はターゲット言語の言語モデルで，$p(f|e)$ は翻訳モデルである．本稿では，IBM Model 2 を用いる．ただし，単語基盤のアライメントモデルであれば，必ずしも IBM Model に限定しないと想定している．

IBM Model 2 は，Python 3.8 において NLTK 3.7 の実装を用いる．学習の繰り返し回数を 20 回に設定する．

3.3 Step 2：ノンリテラル要素の抽出

提案手法の重点である，歌ことばの現代語訳におけるノンリテラル要素の抽出手法について説明する．

各パラレルテキストに対し，Step 1 で学習済みの単語アライメントモデル正方向と逆方向の 2 つで対応づける．歌ことば s について，そのノンリテラル要素を s が出現した個々のパラレルテキストにおける誤り対応として取り扱う．仕組みとして，逆方向モデルでは対象となる歌ことばと対応づけられるが，順方向では対応づけられない訳文における語を誤り対応としてパラレルテキスト一対ずつから探索する．たとえば，パラレルテキスト (S_i, T_i) があり，s が S_i の要素であるとして，順方向モデルでは s が T_i にある要素 t_j に対応づけられ，逆方向では s が T_i にある要素 t_j, t_k, t_l に対応づけられる場合，$\{t_j, t_k, t_l\} \setminus \{t_j\} = \{t_k, t_l\}$ が現代語訳文で s を訳す際に発生したノンリテラル要素と見なす．

3.4 Step 3：ノンリテラル要素の可視化

歌ことばが歌それぞれの訳文にあるノンリテラル要素だけでなく，そのノンリテラル要素の全体像を把握するべく，ネットワーク図で可視化する．ネットワークの作り方について説明する．

歌ことば s について，現代語訳者に共通するノンリテラル要素と各訳者によって異なる要素をグラフ G（式 (3)）に統合して示す．

$$\begin{aligned} G &= (V, E) \\ V &= \{s\} \cup R;\ R = \bigcup_{i=1}^{10000} R_i \\ E &= \{(s,t) | t \in R\} \end{aligned} \tag{3}$$

グラフ G は頂点 V とエッジ E から構成される；V は，歌ことば s の単集合と，その誤り対応語の集合 R の和集合である；$R_1, R_2, R_3, \cdots, R_{10000}$ は 1 から 1000 番目の和歌の 10 人による翻訳それぞれにおいて歌ことば s の誤り対応語の集合を意味する；E は個々の誤り対応関係の集合である．s をさらに訳者ごとにブレークダウンすることで，訳者おのおのの s に関するサブネットワークを連接した形にする．

なお，ノードの媒介中心性でどの誤り対応語がネットワークのハブかを判断する．

3.5 単語アライメントモデルの検証

単語アライメントの精度の検証手法について説明する．

正方向，逆方向の 2 モデルと，双方向のアライメントの積集合について，precision，recall，Alignment Error Rate （AER）[10] を測定する（式 (4)）．

$$\begin{aligned} \text{precision}(A,P) &= \frac{|A \cap P|}{|A|} \\ \text{recall}(A,S) &= \frac{|A \cap S|}{|S|} \\ \text{AER}(A,P,S) &= 1 - \frac{|A \cap P| + |A \cap S|}{|A| + |S|} \end{aligned} \tag{4}$$

A はすべてのアライメントで，S は sure link[*10]で，P は possible link[*11]である．指標の測定に際して，分類語彙表コードが semantic field 番号まで一致する 2 語を S とし，semantic group 番号まで一致する 2 語を P とする（図 2）．AER は低いほど推定精度が高い．

sure link (S)

BG-01-5520-20-0401　梅
BG-01-5520-20-040-A 梅

possible link (P)

BG-01-5520-19-3000　秋萩
BG-01-5520-19-115-A 秋萩

misaligned

BG-01-5520-19-3000　秋萩
BG-08-0061-07-010-A の

図 2 S，P の判定基準：分類語彙表番号の一致の程度でアライメントが sure link か possible link か判定するので，人手による判定より基準が厳しく精度が実際より低下することがある．

3.6 ケース・スタディ

事例研究では，「梅」「松」「桜」「山吹」「女郎花」「菊」の 6 つの典型的な植物の歌ことばを対象にその現代語訳 10 種に見られるノンリテラル要素を抽出しネットワークを描画する．

4. 結果

4.1 単語アライメントモデルの精度

単語アライメントモデルの精度は，precision，recall，および AER で測定した（表 2）．逆方向（target → source）の推定の精度が低く，正方向（source → target）の精度が比較的高かった；正方向と逆方向で推定したアライメントの結合（intersection）の精度が一番高かった．

sure link と possible link の判定規則で精度の値が低くなる可能性がある（考察にて述べる）．

表 2 単語アライメントモデルの精度（%）

	正方向 source → target	逆方向 target → source	結合 intersection
precision	54.98	11.74	69.77
recall	61.27	52.09	56.12
AER	42.39	84.02	37.24

「女郎花」を詠む一例を質的観点から取り上げる（図 3）．この事例では，正方向モデルの推定がもっとも合理的であり，逆方向モデルの推定は逸脱していることがわかった．2 つのモデルの推定したアライメントの積集合（図 3c）は，precision は高いが，正方向モデルにある「10. 我–23. 拙僧」「11. 落つ–31. 堕落」のようなアライメントが除かれていた．結合モデルが保守的であることがわかった．全体的に，低頻度の訳語とのアライメントは不確かな状況が多かった．

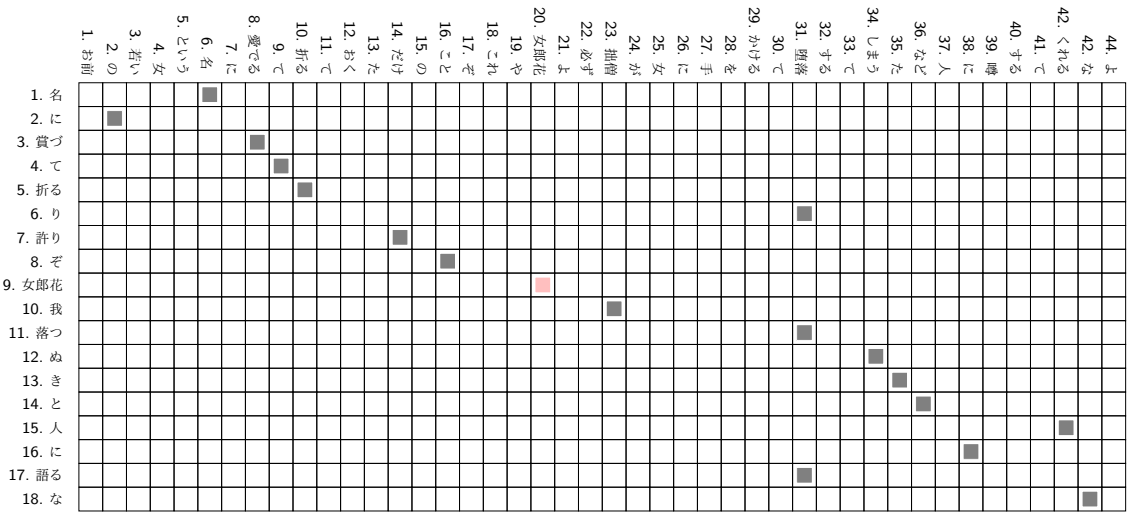
4.2 ノンリテラル要素の事例

図 3 のパラレルテキストから「女郎花」の訳文におけるノンリテラル要素を図 4 に求めた．金子訳 [3] の 226 番歌における「女郎花」のノンリテラル要素が「お前」「女」「よ」と推定されていた．「お前の女という [名]」は確かに「女郎花」に依存しながらも原文には見られない訳語であった．

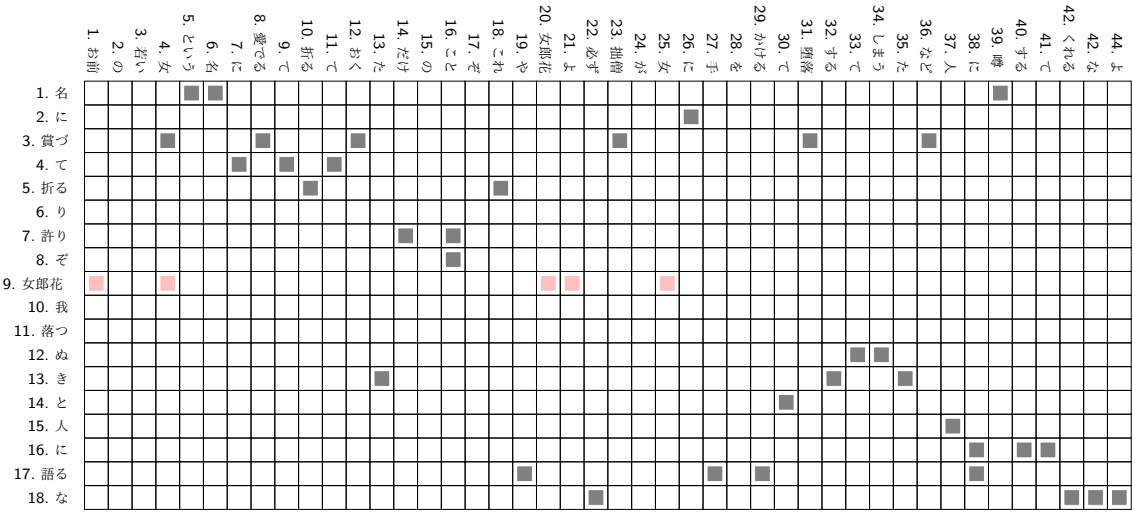
4.3 可視化の結果

可視化の結果，各歌ことばについて現代語訳者が共通して記したノンリテラル要素は，ネットワークのハブとして描画された（図 5）．ハブの性質は，コノテーションに相当する用語（たとえば「女郎花」には「女」「お前」；「梅」には「香」「移り香」；「松」には「待つ」など）以外にも，現代語文を成立させるための機能的な用語（指示語，代名詞，1 語の出現に対して複数回訳された語）が含まれていた．ネットワークには現代語訳のバリエーションがネットワークの周辺ノードとして示された．それぞれの歌ことばに対して各々の訳者が独自に補足した用語が多いことがわかった．

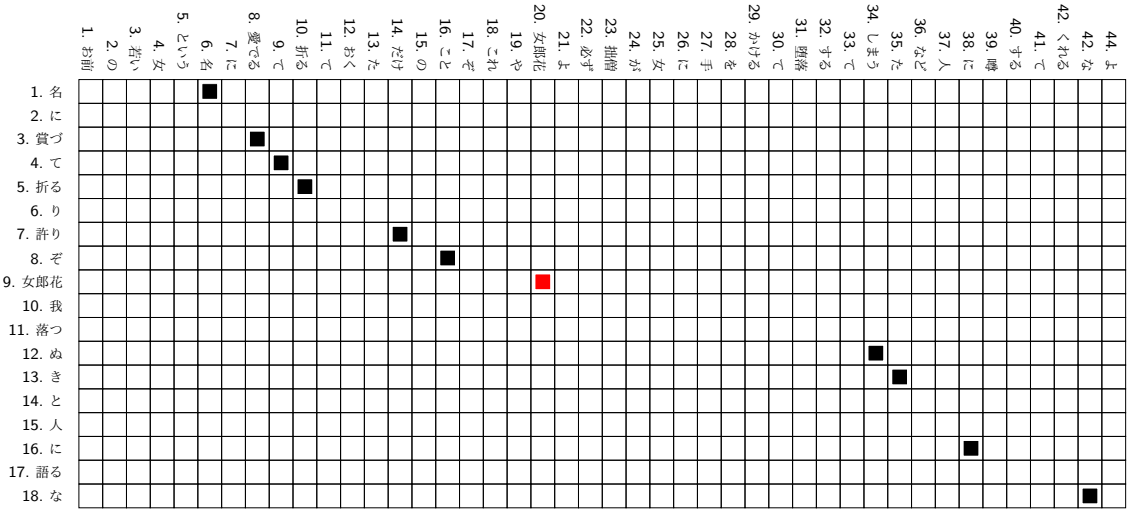
*10 正確なアライメント．
*11 間違いとはいえないアライメント．



(a) 正方向（source → target）アイライメント



(b) 逆方向（target → source）アイライメント



(c) 2 方向の結合（intersection）アイライメント

図 3 金子 [3] による歌番号 226,「女郎花」歌の現代語訳の単語アライメントの結果：色に塗られた格子が推定された歌ことばと現代語訳の対応である；「女郎花」のアイライメントは単方向ではピンク色に，双方向では赤にした．

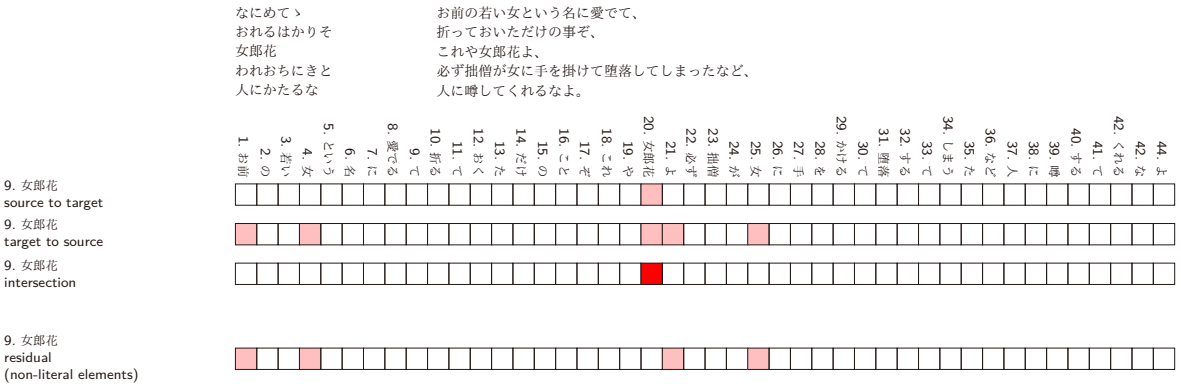


図 4　歌番号 226「女郎花」の金子訳 [3] におけるノンリテラル要素とその抽出過程

5. 考察

5.1 単語アライメントモデルの精度

単語アライメントの精度の計算にあたって，マニュアルに正確なアライメント (S) と可能なアライメント (P) をアノテーションしたデータを用いず，歌ことばと訳語の分類語彙表コードの一致の程度で定義した．そのため，可能な補足の訳語 [*12] は P に含まれず，P が非常に限られた集合になる．その分 recall が下がり AER が上がる．ただし，逆方向モデルの推定の精度が低いわりには，訳語関係にないノンリテラル要素を抽出する手掛かりが提供できると考えられる．

5.2 掛詞の翻訳

ノンリテラル要素の中には，掛詞の複数の意味あいを訳すケースが図 5 に見られた．「松」「菊」はそれぞれ「待つ」「聞く」にかけている．「待つ」「聞く」の 2 語はそれぞれのネットワークにあった．しかし，「待つ」は「松」のネットワークのハブであるのに対して，「聞く」は「菊」のハブではなく 10 人の訳者のうち 3 人にしか補足されなかった．掛詞のすべての意味合いが翻訳されるかどうかは，掛詞としての典型性に関わるであろう．

5.3 上位概念の挿入

ネットワークのハブにある訳語には，歌ことばの上位概念の語，本稿のケーススタディでは「花」が多くとりあげられた．「桜」「菊」「山吹」の現代語訳文に，その上位概念の「花」が多くの訳者により多くの訳文にノンリテラル要素として追加された．ただし，同じ花であるのに，「梅」「女郎花」のネットワークでは，上位概念の「花」がハブにならなかった．つまり，「梅」「女郎花」の訳文には「花」があまり補足されないことであった．その原因は，和歌の原文によく「梅 [の] 花」として梅が詠まれ，その場合「花」が訳文にあるとしても，あくまでも歌ことば「花」の訳語であり「梅」のためにあるノンリテラル要素ではなかったためであろう．「女郎花」の場合は，その漢字表記にすでに「花」の文字が含まれているため，訳者は「花」を補足しなかった可能性がある．

5.4 指示語の挿入

ネットワークのハブに指示語が散見された．「菊」「桜」「梅」に関しては指示語「この」がノンリテラル要素のハブであった．「女郎花」に関しては指示語「あの」がハブであった．花に関連することばによって現代語訳に使われる指示語のバリエーションが見られた．ほかの花と異なり，「女郎花」は知覚的に遠い存在（思い人のメタファー）の認識が専門家・訳者にあるといえる．

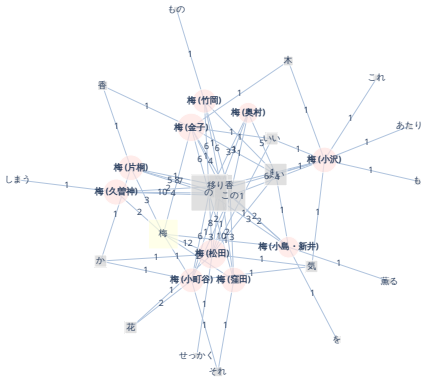
5.5 訳者間に共通したコノテーション

ハブの中に前述していない語はそれぞれの歌ことばのコノテーションに相当するノンリテラル要素であった．例えば，「女郎花」に「女」，「梅」に「移り香」「香」，「松」に「変わる」，「菊」に「盛り」などが見られた．これらの意味合いは，それぞれ和歌の原文には明示的に詠まれずに平安期の歌人のコミュニティにわかるものと考えられる．現代の読者がわかるように多くの専門家がそれらの意味合いを明示的に訳している．一方で，ハブにない周縁に分布するノードは，訳者個人個人が補足したものであった．ノンリテラル要素の個人差が伺えた．

一部掛詞の翻訳もコノテーションと捉えてよいが，上位概念，指示語のような機能的なノンリテラル要素は，コノテーションとして不適切であった．今後データから機能的要素を取り除いた上で，アライメントを推定するなど，前処理の工夫をする余地があることがわかった．

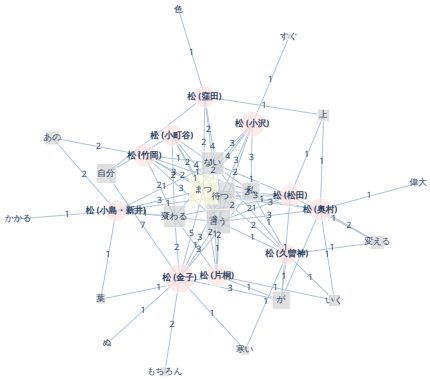
[*12] 品詞が異ったなり，空へのアライメントなりで分類コードがマッチングできない．

BG-01-5520-20-0401 梅



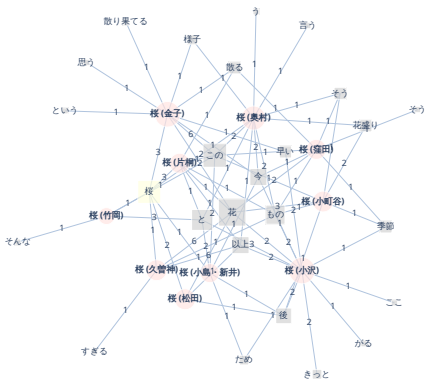
top-5 betweenness centrality: **移り香: 64.11; この: 64.11; の: 64.11; 梅: 24.51; よい: 24.04**

BG-01-5520-34-0301 松



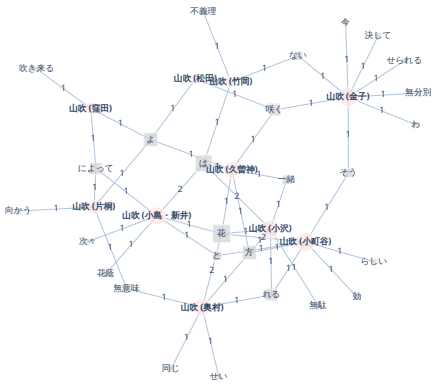
top-5 betweenness centrality: **待つ: 61.91; まつ: 50.50; 言う: 43.42; ない: 25.37; 変わる: 21.77**

BG-01-5520-20-1101 桜



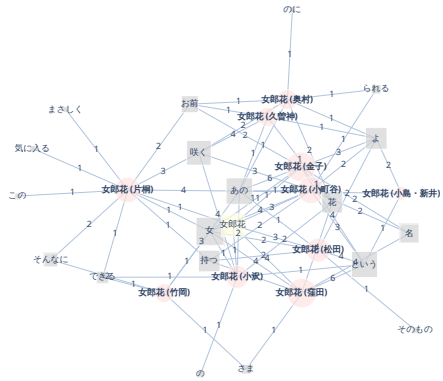
top-5 betweenness centrality: **花: 61.98; この: 60.35; 桜: 59.96; 今: 41.85; もの: 40.37**

BG-01-5520-20-1700 山吹



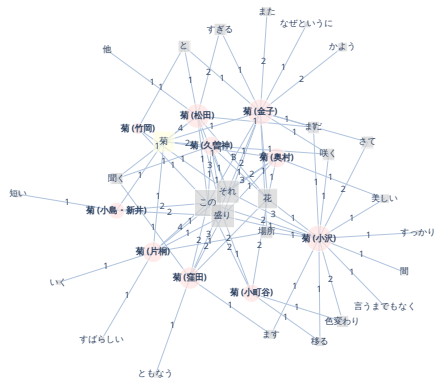
top-5 betweenness centrality: **花: 113.24; 咲く: 103.60; は: 94.08; よ: 92.72; そう: 70.68**

BG-01-5520-05-0106 女郎花



top-5 betweenness centrality: **あの: 71.17; 女: 36.23; 女郎花: 24.45; 咲く: 20.53; よ: 17.10**

BG-01-5520-03-0600 菊



top-5 betweenness centrality: **この: 105.46; 盛り: 87.02; それ: 62.81; 花: 40.51; 菊: 28.67**

図 5 植物類の歌ことば 6 つについて，現代語訳で補足したノンリテラル要素のネットワーク可視化：ノードは対象となる歌ことば（四角いノード）とその現代語訳におけるノンリテラル要素（丸いノード）である．エッジは歌ことばとそれぞれのノンリテラル要素とのアライメント関係である；ノードの大きさ，エッジ上の数字は，ノンリテラル要素が対象とする歌ことばの補足と推定された回数を示す；図の下に歌ことばノードを除いて媒介中心性上位 5 の訳語ノードを記述する．

5.6 総合討議

提案手法による和歌–現代語訳対からのノンリテラル要素候補の抽出は，先行研究 [15], [16] に比べ，さらに絞られただけでなく，ネットワークを構成するノンリテラル要素がどの和歌の翻訳の中で追加されたかという中間過程も確認できた．しかし，コノテーションに該当しない機能的な用語（指示語，代名詞，1 語の出現に対して複数回訳された語）も多く含まれていた．翻訳の補足を抽出したとはいえ，コノテーションの抽出手法としては，機能的な品詞類と同語を除外するなど前処理に工夫を凝らす必要があると考えられる．

単語アライメントモデルの精度については，人手による作業に比べては低いものの，精度測定の条件にしては不十分な精度ではなかった．古代語の分析を現代人の主観を交えない研究の趣旨に貢献できたといえる．

ネットワークの周辺に散見されたノンリテラル要素は，各々の訳者間で認識が異なることが起因しているのであれば，コミュニケーション空間におけるエンコード（言語化）の個人差研究として今後展開できよう．さらに個人差を越えて，異なる言語変種のパラレルコーパスから「同じことの異なる言い方（語彙変数）」を検出・収集し，単語アライメントの技術を（認知）社会言語学の研究に応用できるであろう．

方法論の拡張性に関しては，本稿で用いた IBM Model に限らず，さらに発展した単語アライメントモデルの適用も可能である．使用テキストについて，日本人の翻訳のみならず，多言語の翻訳を用いることもできる．

6. おわりに

図 1 にあるように，(a)［歌人が詠む →（和歌という信号）← 現代語訳者が読む］で明示的に言語化する必要のないもの（コノテーション）を (b)［現代語訳者が書く →（現代語訳という信号）← 一般読者が読む］でも正しく伝わるように，訳者が (a) での信号（和歌）を (b) での信号（現代語訳）へと書き換える．その書き換えとは，(a) における言外の物事を (b) で明示することである．訳者が歌ことばのコノテーション（の候補）を現代語訳として明示する行動（ノンリテラルな訳語の挿入）を，本稿ではパラレルテキストに対する単語アライメントの誤り対応として抽出した．さらに，訳者間で共通する・共通しない認識をネットワーク形式で可視化した．

その結果，ノンリテラルな訳語は，対象となる歌ことばのコノテーションのみならず，現代語の文として成立させるために必要な機能的な要素も含まれていた．そのため，コノテーションの抽出手法として，機能語を除外するなど前処理を工夫する必要がある．また，植物類の歌ことばの事例では，訳者間で共通したノンリテラルな訳語が限定され，コノテーションとして補足される表現の個人差が大きいことが観測できた．

謝辞　本研究は JST 科学技術イノベーション創出に向けた大学フェローシップ創設事業 JPMJFS2112 の支援を受けた．

参考文献

[1] Brown, P. E., Pietra, V. J. D., Pietra, S. A. D. and Mercer, R. L.: The Mathematics of Statistical Machine Translation: Parameter Estimation, *Computational Linguistics*, Vol. 19, No. 2, pp. 263–311 (1993).
[2] Hodošček, B. and Yamamoto, H.: Development of Datasets of the Hachidaishū and Tools for the Understanding of the Characteristics and Historical Evolution of Classical Japanese Poetic Vocabulary, *Digital Humanities 2022 Conference Abstracts*, Tokyo, The University of Tokyo, pp. 647–648 (2022).
[3] 金子元臣：古今和歌集評釈: 昭和新版，明治書院，東京 (1933).
[4] 片桐洋一：古今和歌集全評釈，Vol. 上中下，講談社，東京 (1998).
[5] 小島憲之，新井栄蔵：古今和歌集，岩波書店，東京 (1989).
[6] 小町谷照彦：古今和歌集：現代語訳対照，旺文社，東京 (1982).
[7] 窪田空穂：古今和歌集評釈，Vol. 上中下，東京堂，東京 (1960).
[8] 久曽神昇：古今和歌集 全注釈，講談社，東京 (1979).
[9] 松田武夫：新釈古今和歌集，Vol. 上下，風間書房，東京 (1968).
[10] Och, F. J. and Ney, H.: A Systematic Comparison of Various Statistical Alignment Models, *Computational Linguistics*, Vol. 29, No. 1, pp. 19–51 (online), DOI: 10.1162/089120103321337421 (2003).
[11] 奥村恆哉：古今和歌集，新潮社，東京 (1978).
[12] 小沢正夫：古今和歌集，小学館，東京 (1971).
[13] Schramm, W. L.: *The Process and Effects of Mass Communication.*, University of Illinois Press, Urbana (1954).
[14] 竹岡正夫：古今和歌集全評釈: 古注七種集成，Vol. 上下，右文書院，東京 (1976).
[15] Yamamoto, H.: A Mathematical Analysis of the Connotations of Classical Japanese Poetic Vocabulary, PhD Thesis, The Austrilian National University, Canberra (2005).
[16] 山元啓史：歌ことばの可視化とコノテーションの抽出－グラフによる共出現パターンの作り方－，じんもんこん 2006 論文集， Vol. 2006, pp. 21–28 (2006).
[17] Yamamoto, H. and Hodošček, B.: An Analysis of the Differences Between Classical and Contemporary Poetic Vocabulary of the Kokinshu, *The 9th Conference of Japanese Association for Digital Humanities (JADH2019) "Localization in Global DH"*, pp. 68–71 (2019).