

直接調停要約自動生成システム HERMeS の 実装と性能評価

中野 正寛[†] 洪木 英潔[‡] 宮崎 林太郎[†] 石下 円香[‡] 永井 隆広[†] 森 辰則[‡]
[†]横浜国立大学 大学院 環境情報学府 [‡]横浜国立大学 大学院 環境情報研究院
 E-mail: {nakano,shib,rintaro,ishioroshi,nagadon,mori}@forest.eis.ynu.ac.jp

1 はじめに

Web 上の情報は、ブロードバンド化の進展やブログ等の普及に伴い、爆発的に増加し続けている。これらの中には、出所が不確かな情報や利用者に不利益をもたらす情報などが含まれており、信頼できる情報を利用者が容易に得るための技術に対する要望が高まっている。しかし、情報の内容の真偽や正確性を検証することは困難であるので、我々は、利用者の信憑性判断を支援する技術に関する研究を行っている [1, 2, 3]。

現在、ある情報の信憑性を Web のみを情報源として判断しようとした場合、検索エンジンにより上位にランキングされた文書集合を読んで判断することが多い。しかし、その文書集合に、判断の対象となる情報を肯定する文書と否定する文書の両方が含まれている場合、その対立関係をどのように読み解くべきかに関する手がかりを検索エンジンは示さない。ここでの対立関係の読み解き方とは、例えば、一方の内容が間違っているのか、それとも、両方の内容が正しく両立できるのか、といった点に関する可能性の示唆であり、両立できるのであれば、何故対立しているようにみえるのかに関する解説を提示することが望ましい。

以上のことから、我々は、一見対立しているようにみえるが、実際はある条件や状況の下で互いの内容が両立できる関係を**擬似対立**と呼び、擬似対立を読み解くための手掛かりを簡潔に提示する**調停要約**を提案している [1]。我々は、利用者が信憑性を判断したい言明¹（以降、**着目言明**）を入力した際に調停要約²を作成する、「調停要約自動生成システム HERMeS (Helpful Explanation Representing Mediatory Summary)」の作成を進めており、本稿では調停要約の自動生成手法についての提案と評価実験を行う。

本稿の構成は以下の通りである。2 章では、研究の背景と、調停要約に付いて説明する。3 章では、提案手法のアルゴリズムを述べる。4 章では、評価実験を行い、その結果を考察する。5 章では、まとめを述べる。

2 研究の背景

2.1 関連研究

情報信憑性の判断を支援する技術には幾つかのアプローチが考えられる。まず、利用者が着目するトピックに対して、言明間に存在する根拠情報や矛盾・対立情報といった論理的关系を提示することで、利用者が情報を多角的に俯瞰することを支援する技術がある。河原

¹ 本稿では、主観的な意見や評価だけでなく、疑問の表明や客観的事実の記述を含めたテキスト情報を広く**言明**と呼ぶこととする。

² 文献 [1] では、調停要約として直接調停要約と間接調停要約の 2 種類が定義されているが、本稿では直接調停要約を対象としており、以後、直接調停要約を単に調停要約と記す。

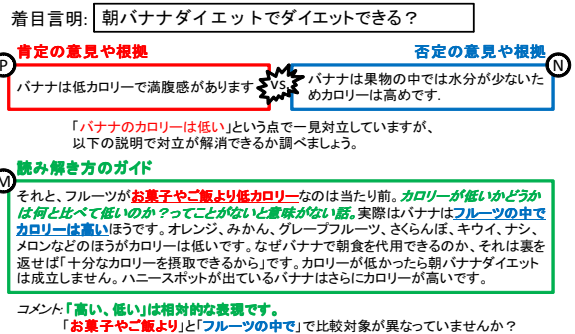


図 1: 調停要約の例

ら [4] は、述語項構造に基づき、あるトピックに関する主要・対立表現を俯瞰的に提示するシステム WISDOM を開発しており、村上ら [5] は、論理的关系の種類を拡張した言論マップの生成課題を論じている。

我々は、これらの技術と並んで、擬似対立や、その読み解き方を提示する技術も重要であると考え、調停要約に関する研究に取り組んでいる。言明間の論理的关系の把握が容易になるような要約を利用者に提示できれば、着目言明に関連する言明群の全体像が把握しやすくなると考えられるので、文献 [4, 5] の技術とは相補的な関係であるといえる。

調停要約は複数文書要約の一つである。その中でも、橋本ら [6] の研究のような、要約対象文書群から「まとめ文章」を取り出すことにより要約する手法に属する。文献 [6] では、新聞記事を対象にしているので、複数文書の記述内容に齟齬があることは想定せずに、複数記事の内容をそのまままとめることが目的である。一方で調停要約では、まず、様々な立場の人物や組織が互いに対立した主張をしているようにみえる記述を含む文書集合を要約対象にしている点が異なる。さらに、得られる要約の中に、疑似対立とその読み解き方が含まれるようにすることで、情報信憑性の判断に寄与することを目的としている。

2.2 調停要約

我々が目標とする調停要約の例を図 1 に示す。着目言明に関して、対立関係にあるようにみえる記述を (P) と (N) のボックス内に表示している。Web 上には、こういった対立関係について、それらが両立可能であることを示した記述が存在することがあり、そのような記述をパッセージ単位で抜粋して提示するというのが調停要約の基本的な考え方である。図 1 では、(M) のボックス内が調停要約として抜粋された記述である。

我々は、信憑性判断の支援のための要約コーパスを手で作成し分析を行った。その結果、作成された要約には疑似対立や調停要約に相当する記述が多く含ま

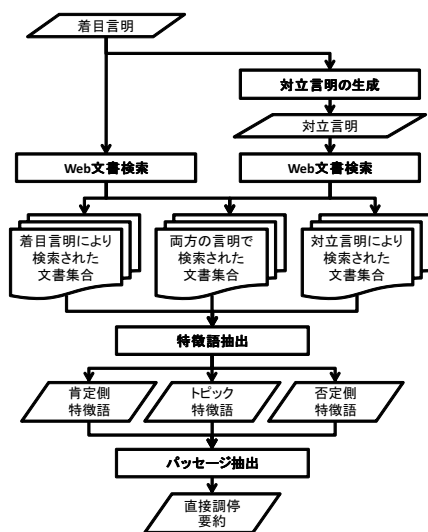


図 2: 提案手法の全体の流れ

ることが分かった [2]. それゆえ、我々は、信憑性判断の支援において調停要約が重要であると考え、分析を行った結果、調停要約として適切な記述には以下の特徴があることが分かり、それらの特徴を用いて調停要約として適切なパッセージの抽出を行うこととした。

第一の特徴は、着目言明との関連性である。第二の特徴は、公平性、すなわち、着目言明を肯定および否定する意見や根拠等の両方に言及していることである。調停要約は、擬似対立を読み解くための手がかりとなる記述であるので、両方の主張が含まれるパッセージはより適切であると考えられる。第三の特徴は、上記の特徴を表す単語の密集度である。高い密集度をもつパッセージは、簡潔な要約としての妥当性に加えて、両方の意見や根拠等を対比的に記述していることが多く、調停要約としてより適切であると考えられる。

3 提案手法

3.1 パッセージ抽出に基づく調停要約生成タスク

調停要約生成タスクでは、利用者により与えられる着目言明を入力とする。通常の要約生成タスクでは、入力として要約対象となる文書集合が与えられるが、調停要約生成タスクでは、与えられた着目言明に関連した文書群を Web 上から検索して、要約対象とする。出力として、着目言明に関連する疑似対立の読み解き方の手がかりを示すパッセージを抜粋して提示する。着目言明に関連する疑似対立は 1 つであるとは限らないので、利用者には複数の調停要約を提示することが前提となる。

3.2 提案手法の概要と全体の流れ

我々は、2.2 節で述べた特徴をもつパッセージを調停要約として抽出するために、以下の 3 種類の特徴語に基づく手法を提案する。

まず、着目言明との関連度を近似的に求めるために、着目言明のトピックとの関連性が高い語を**トピック特徴語**として定義する。本稿では、着目言明中の内容語をトピック特徴語とした³。次に、着目言明を肯定また

³ トピック特徴語は論理的に着目言明中の単語に限定されるものではない。着目言明以外の語句への拡張は今後の課題である。

は否定する意見や根拠等と言及しているかを近似的に求めるために、双方の意見や根拠に現れやすい単語を**肯定側特徴語**および**否定側特徴語**としてそれぞれ定義する。肯定側特徴語と否定側特徴語は対となる語であると考えられるが、着目言明中に含まれているとは限らず、また、必ずしも一般的な対義語に限定されない⁴。それゆえ、これらは着目言明以外から決定する必要がある。本稿では、要約対象となる文書集合を着目言明に肯定的な文書集合と否定的な文書集合に分類して、双方の意見や根拠に現れやすい単語を抽出する。

図 2 に提案手法の全体の流れを示す。入力として着目言明が与えられると、まず、着目言明と反対の極性をもつ対立言明の生成を行う。次に、着目言明と対立言明を用いて Web 文書集合の検索を行う。検索された文書集合は、肯定側特徴語および否定側特徴語を抽出するために用いる。抽出された 3 種類の特徴語を用いて、2.2 節で述べた特徴に基づき、検索された Web 文書中の各パッセージについて直接調停としてのスコアを計算し、スコアに応じてパッセージを出力する。本章の次節以降で、上記の各段階を詳しく説明する。

3.3 対立言明の生成と Web 文書の検索

対立言明は、対義語辞書を用いて、着目言明中の 1 語をその対義語で置換することにより生成する。また、着目言明中の置換された語を肯定側特徴語、置換された語に対応する語を否定側特徴語として保持する。

次に、着目言明と生成された対立言明をクエリとして、自然文をクエリとして使用できる TSUBAKI[7]により Web 文書の検索を行う。検索された文書は以下の 3 種類の文書集合に分類する。第一の集合 D_{query} は、着目言明により検索されたが対立言明により検索されなかった文書の集合、第二の集合 $D_{inverse}$ は、対立言明により検索されたが着目言明により検索されなかった文書の集合、第三の集合 D_{both} は、着目言明と対立言明の両方で検索された文書の集合である。

TSUBAKI には、クエリ中の最後の用言について、その反義・否定表現を加えて検索する機能があり、その機能を用いた場合に着目言明により検索された文書を、対立言明により検索された文書とみなすこととした。また同様に、その機能を用いた場合に対立言明により検索された文書を、着目言明により検索された文書とみなすこととした。これにより、対義語辞書で対立言明が生成されなかった場合でも、着目言明と対立する内容を含む文書を検索することが可能となる。

3.4 特徴語の抽出

肯定側特徴語と否定側特徴語は、検索された Web 文書から抽出する。最初に、単語 w における肯定側スコア $sc_{POS}(w)$ と否定側スコア $sc_{NEG}(w)$ を、式 (1) と式 (2) によりそれぞれ計算する。

$$sc_{POS}(w) = \frac{df(w, D_{query}) \cdot tf(w)}{df(w, D_{inverse}) + 1} \quad (1)$$

$$sc_{NEG}(w) = \frac{df(w, D_{inverse}) \cdot tf(w)}{df(w, D_{query}) + 1} \quad (2)$$

ここで $tf(w)$ は検索された全ての文書における単語 w の頻度であり、 $df(w, D)$ は文書集合 D において w を

⁴ 例えば、「キシリトールは虫歯に効果がある」という言明に着目した場合には、虫歯の「予防」には効果があるが「治療」には効果がないといったように「予防」と「治療」が対となる単語となる。

含んでいる文書の数である。

検索された全文書に満遍なく出現する語は、 $tf(w)$ が高いので $sc_{POS}(w)$ と $sc_{NEG}(w)$ の両方が高い値となりやすいが、肯定側特徴語や否定側特徴語として抽出すべきではない。それゆえ、抽出には $sc_{POS}(w)$ と $sc_{NEG}(w)$ の値にそれぞれ基づいた順位 $rank_{POS}(w)$ と $rank_{NEG}(w)$ を用いることとした。まず、全ての単語を $tf()$ の値が大きい順に順位付けし、上位 C_{rank} 語を、抽出する特徴語の候補とする。その候補の中で、 $rank_{POS}(w)$ と $rank_{NEG}(w)$ の差が閾値 C_{dif} よりも大きい単語を肯定側特徴語または否定側特徴語とし、その他の単語は特徴語とみなさないようにした。最後に、3.3 節で対立言明を生成する際に保持した肯定側および否定側の特徴語集合を、抽出された肯定側および否定側の特徴語集合にそれぞれ加える。

トピック特徴語集合は、肯定側および否定側の特徴語集合に含まれない着目言明中の単語の集合とした。

3.5 パッセージ抽出とランキング

パッセージは以下の処理で抽出・ランキングする。

第一段階の処理では、調停要約として適切ではない文を認識する。本稿では、1 以上の動詞と 2 以上の名詞、または、2 以上の動詞と 1 以上の名詞をもつ文を十分な文と定義し、これを満たさない文を不十分な文とした。また、文末が「...」という省略表現で終わっている文を省略された文とし、以上の 2 種類を調停要約として適切ではない文とした。

第二段階の処理では、文のスコアを計算する。 KW を、トピック特徴語集合、肯定側特徴語集合、否定側特徴語集合の和集合とした場合、文 s の基本スコア $sc_{BAS}(s)$ は式 (3) によって計算する。

$$sc_{BAS}(s) = \frac{\sum_{w \in KW} appear(w, s)}{|KW|} \quad (3)$$

ここで $appear(w, s)$ は単語 w が文 s 中に現れていれば 1、そうでなければ 0 を返す関数である。もしも、 s が不十分もしくは省略された文であった場合、 $sc_{BAS}(s)$ にペナルティとして 1 より小さい定数 $P_{insufficient}$ もしくは 1 より小さい定数 $P_{incomplement}$ を乗じる。また、2.2 節で述べた公平性の観点から、文 s がトピック特徴語に言及している⁵ ことに加えて、肯定側特徴語または否定側特徴語のどちらかもしくは両方に言及している場合、 $sc_{BAS}(s)$ に 1 より大きい定数 R_{either} もしくは R_{either} より大きい定数 R_{both} を乗じる。

第三段階の処理では、パッセージの断片化を防ぐために文スコアを平滑化する。窓幅が 5 のハニング窓を用いて、式 (4) に従って、文書中の i 番目の文の平滑化スコア $sc_{SMO}(s_i)$ を計算する。

$$sc_{SMO}(s_i) = \sum_{j=-\frac{L}{2}}^{\frac{L}{2}} (sc_{BAS}(s_{i+j}) \cdot hf(j)) \quad (4)$$

$$hf(k) = 0.5 + 0.5 \cos 2\pi \frac{k}{L} \quad (5)$$

パッセージ中の前後の文と適切につながること、不十分な文は読み手に有益な情報を伝える可能性があるが、省略された文の場合は基本的にそうはならない。それゆえ、不十分な文の $sc_{SMO}(s)$ は平滑化を行うが、

⁵ 本稿での「特徴語に言及している」とは、例えば、複数ある特徴語の内の少なくとも 1 語が現れている状態のことを意味する。

省略された文は平滑化を行わないこととした。文 s を中心とした窓内の文列が、トピック特徴語、肯定側特徴語、否定側特徴語の全てに言及している場合、公平性の観点から s は調停要約の一部である可能性が高いとして $sc_{SMO}(s)$ に 1 より大きい定数 R_{smooth} を乗じる。

第四段階の処理では、パッセージの抽出とランキングをする。文書中における $sc_{SMO}(s)$ の最大値の $\frac{1}{3}$ 以上の $sc_{SMO}(s)$ をもつ文の連続を 1 つのパッセージとして抽出する。パッセージ p のスコア $sc_{PAS}(p)$ は、パッセージ中の文の $sc_{SMO}(s)$ の最大値とする。 p 中の文列が、トピック特徴語、肯定側特徴語、否定側特徴語全てに言及している場合、公平性の観点から p は調停要約である可能性が高いとして $sc_{PAS}(p)$ の値に 1 より大きい定数 $R_{passage}$ を乗じる。また、簡潔性という観点から、抽出されたパッセージの長さが目標とする長さ C_{length} に近いほど上位にランキングされるべきである。それゆえ、パッセージの最終スコア $sc_{FIN}(p)$ は式 (6) に従って計算される。

$$sc_{FIN}(p) = \exp(sc_{PAS}(p) - C_{error} \cdot er(p)) \quad (6)$$

$$er(p) = |C_{length} - nc(p)| \quad (7)$$

ここで $nc(p)$ はパッセージ p の文字長、 C_{error} は目標とする長さとの誤差の影響に関する係数である。

4 評価実験

4.1 目的と手順

調停要約という新しい要約の概念に基づく要約手法は今までに存在しないので、提案手法と直接比較可能な従来手法は存在しない。それゆえ、以下の 3 手法と提案手法を比較することとした。

第一の手法 $KWtf$ は、3 章で述べた 3 種類の特徴語を用いずに提案手法と同じ流れで要約する手法である。第二の手法 Lin_{TSU} は、TSUBAKI による検索ランキング上位の文書を既存の要約モジュールを用いて要約する手法である。要約モジュールは、文書中の単語の出現頻度と接続頻度に基づいて特徴的な語句を含む文を抽出する `Lingua::JA::Summarize::Extract`⁶ を用いた。第三の手法 Lin_{scFIN} は、パッセージスコア $sc_{FIN}()$ が上位のパッセージを含む文書を Lin_{TSU} と同じ要約モジュールを用いて要約する手法である。

提案手法の特徴の一つは、公平性を考慮するために、 $tf()$ による上位 C_{rank} 語を、2 つの特徴語および非特徴語に分類することである。 $KWtf$ では、この分類を行わずに上位 C_{rank} 語を全て、極性を持たない特徴語とする。提案手法と $KWtf$ との比較により、特徴語の分類により公平性を考慮することの有効性を調査する。

Lin_{TSU} と提案手法の比較は、文書中の特徴的な語句を含む文を抽出する一般的な目的の要約と、我々が目的とする調停要約との違いを調査するために行う。

提案手法のもう一つの特徴は、文書検索時のスコアと無関係にパッセージのランキングを行うことである。それゆえ、提案手法により上位にランキングされたパッセージを含む文書集合が、 Lin_{TSU} の要約対象となる文書集合と異なる可能性がある。提案手法と同じ文書集合を要約対象とする Lin_{scFIN} と提案手法との比較により、提案手法が調停要約として適切なパッセージを抽出しているかを調査する。

⁶ <http://search.cpan.org/~yappo/Lingua-JA-Summarize-Extract-0.02/>

表 1: 調停要約としての適切性に関する全体の適合率

	Top 5	Top 10	Top 15	Top 20
Proposed	27.2%	24.4%	25.6%	23.0%
KWtf	5.6%	9.6%	10.1%	11.2%
Lin _{TSU}	19.2%	18.4%	15.7%	14.0%
Lin _{scFIN}	18.4%	19.6%	18.7%	18.2%

調停要約は新しい概念であるので、調停要約としての適切性を評価するための手法は未だ確立されていない。要約の評価手法として ROUGE[8] が一般的であるが、N-gram による再現度のスコア付けは 2.2 節で述べた公平性を考慮することが困難であるので、調停要約の評価手法として適切ではない。それゆえ、各手法により出力された上位 20 件の要約を対象に、対立しているようにみえる二言明が両立可能となる状況を読み手が理解できるような記述が含まれているか否かを二値判断で評価することとした。

調停要約では、簡潔な要約であることよりも、両立可能となる状況を読み手が理解するのに十分な情報を提示することの方が重要であると考えられる。それゆえ、要約長に関する制限は設けなかった。また、評価は両立可能となる状況の提示という観点からのみとし、簡潔性や冗長性による影響がないようにした。評価は、「割り箸は環境に良い」のように主観的でない判断が可能な 25 の言明について、第一著者が行った。

我々は、出力結果の上位に少数でも調停要約として適切なパッセージが提示できれば良いので、調停要約では再現率よりも適合率が重要だと考える。また、再現率を求めるために、調停要約として適切なパッセージが Web 上にいくつ存在するかを調べることが現実的に不可能であること、実際の利用を想定すると、利用者は少数の要約しか目を通さないであろうことを考慮した結果、本稿では、評価尺度として適合率を用いた。

提案手法のパラメータとして、予備実験の結果から、 $P_{insufficient}$, $P_{incomplete}$, R_{either} , R_{both} , R_{smooth} , $R_{passage}$, C_{rank} , C_{dif} , C_{length} , C_{error} の値を、それぞれ、0.5, 0, 2, 3, 2, 3, 100, 20, 300, 0.02 とし、Lin_{TSU} と Lin_{scFIN} では要約長を 500 文字以内とした。TSUBAKI による文書検索において、クエリごとに検索する文書数は 100 とした。

4.2 結果と考察

表 1 に、各手法の調停要約としての適切性に関する適合率をそれぞれ示す。表 1 から、全体として提案手法が最も良い結果となったことが分かる。検索された文書の全てが調停要約として適切なパッセージを含んでいるわけではないので、達成可能な適合率が必ずしも 100% になるとは限らない。それゆえ、調停要約として適切なパッセージをより上位にランキングする手法が有効な手法と考えることができる。提案手法は、他の手法と比較してより上位に調停要約として適切なパッセージがランキングされることが多かった。

提案手法の適合率が KWtf を大きく上回っていることから、頻出語を 3 種類の特徴語に分類することの有効性が確認できる。それゆえ、2.2 節で述べた公平性が、適切な調停要約を生成する上で重要な役割を果たしているといえる。また、提案手法と同じ文書集合を要約対象としている Lin_{scFIN} の適合率が提案手法に及ばないことから、提案手法は既存の要約手法より調停要約の生成手法として適切であることが確認できる。

いくつかの言明では、提案手法が Lin_{TSU} よりも上位に調停要約として適切なパッセージをランキングできなかった。提案手法では文書検索時のスコアと無関

係にパッセージのランキングを行っていることから、文書検索時のスコアを考慮することで調停要約としての適切性に関する指標を改善できる可能性がある。

今回使用した 25 言明の中では、1 言明のみ、全ての手法において、調停要約として適切であると判断された要約が存在しなかった。この言明については、検索文書集合中に、調停要約として適切であるパッセージが存在しなかった可能性がある。

5 おわりに

本稿では、Web 上の情報信憑性に関する利用者の判断を支援するための直接調停要約の自動生成手法を提案して、性能の評価を行った。その結果、既存の要約手法と比較して、提案手法が調停要約に適した要約を生成できたことを確認した。

今後の課題として、2.2 節で述べた、関連性、公平性、密集度の他にも、調停要約としての適切性を判断する指標がないかを検討したい。

謝辞

本研究は、独立行政法人情報通信研究機構の委託研究「電気通信サービスにおける情報信憑性検証技術に関する研究開発」プロジェクトの成果である。

参考文献

- [1] K. Kaneko, H. Shibuki, M. Nakano, R. Miyazaki, M. Ishioroshi, and T. Mori, Mediatory Summary Generation: Summary-Passage Extraction for Information Credibility on the Web, *Proc. of the 23rd Pacific Asia Conference on Language, Information and Computation*, pp.240–249, 2009.
- [2] M. Nakano, H. Shibuki, R. Miyazaki, M. Ishioroshi, K. Kaneko, and T. Mori, Construction of Text Summarization Corpus for the Credibility of Information on the Web, *Proc. of the 7th Language Resources and Evaluation Conference*, pp.3125–3131, 2010.
- [3] H. Shibuki, T. Nagai, M. Nakano, R. Miyazaki, M. Ishioroshi, and T. Mori, A Method for Automatically Generating a Mediatory Summary to Verify Credibility of Information on the Web, *Proc. of the 23rd International Conference on Computational Linguistics*, pp.1140–1148, 2010.
- [4] 河原大輔, 黒橋禎夫, 乾健太郎, 主要・対立表現の俯瞰的要約ウェブの情報信頼性分析に向けて, 信学技法 Vol.108 No.141, 2008-NL-186, 2008.
- [5] 村上浩司, 松吉俊, 隅田飛鳥, 森田啓, 佐尾ちとせ, 増田祥子, 松本裕治, 乾健太郎, 言論マップ生成課題: 言説間の類似・対立の構造を捉えるために, 情報処理学会研究報告, 自然言語理解・言語理解とコミュニケーション合同研究会, 信学技報 Vol.108 No.141, 2008-NL-186, pp.55-60, 2008.
- [6] 橋本力, 奥村学, 島津明, 複数記事要約のためのサマリパッセージの抽出, 言語処理学会第 7 回年次大会発表論文集, pp.285-288, 2001.
- [7] K. Shinzato, T. Shibata, D. Kawahara, C. Hashimoto, and S. Kurohashi, TSUBAKI: An Open Search Engine Infrastructure for Developing New Information Access Methodology, *Proc. of the Third International Joint Conference on Natural Language Processing*, pp.189–196, 2008.
- [8] C. Lin and E. Hovy, Automatic Evaluation of Summaries Using N-gram Co-Occurrence Statistics, *Proc. of the Human Language Technology Conference 2003*, pp.71–78, 2003.