文章の光と影：可読性向上のための文章作成支援

西原陽子 [1,2]　砂山渡 [3]　谷内田正彦 [1]
[1] 大阪大学大学院基礎工学研究科　[2] 日本学術振興会
[3] 広島市立大学大学院 情報科学研究科
〒560-8531　大阪府豊中市待兼山町 1-3
yoko@yachi-lab.sys.es.osaka-u.ac.jp

Abstract: 文章は多くの人が使用できるコミュニケーションツールの 1 つだが，自らの意図を正しく伝え，過不足なく説明がなされた文章を書くことは難しい．大多数の人が書き手の意図を理解できる，説明が足りている文章を書くための支援が必要とされている．本研究では，文章において，説明がきちんとなされた部分を光，説明が不足している部分を影とし，文章の光と影を評価するシステムを提案する．提案システムではテーマを表す単語やテーマに関係する単語を特定し，それらを光源として文章において光で照らされている部分とそうでない部分を評価し，光と影の分布状態を視覚インタフェース上に出力する．また，光の部分に頻出する単語，影の部分に頻出する単語を評価し，それぞれの単語のリストも出力する．光と影の分布状態と 2 種類のリストを用いることによって，ユーザの文章作成の支援を図る．

1．はじめに

書き言葉，話し言葉などは多くの人が使えるコミュニケーション手段の一つである．手軽に使えることから日常会話，メール作成，文書作成など，大部分において言葉が用いられている．言葉は簡単に使えるためか，意識して正しく使おうと考えられることは少ない．そのため，敬語の使い方や「ら抜き言葉」などの表面的な誤りだけでなく，説明が不足していて，自らの意図を正しく伝えられない事態が頻繁に起こっている．対面の会話であれば，意図が伝わらなかったと分かったときに，付加して説明を加えることによって，自分の一番伝えたい意図に近いところまで修正できるが，新聞記事，コラム，小説，小論文，論文など書かれた文章は，説明を加えることが難しい欠点がある．そのため，大多数の人が書き手の意図を理解できる，説明が足りている文章を書く必要がある．

そこで，本研究では，文章において，説明がきちんとなされた部分を光，説明が不足している部分を影とし，図 1 のように文章の光と影を評価するシステムを提案する．多くの写真がある中で綺麗だと選ばれる写真は，光と影の対比がはっきりととらえられていることが多い．これは文章を写真に見立てたときにも同じことが言え，良い文章ならば説明がきちんとなされた光の部分，曖昧に含みをもたせた影の部分がバランスよく含まれていると考えられる．例えば，原著論文やプロジェクトの報告書などは，説明に不足があってはならないため，文章全般に渡って光が当たった状態になっていると考えられ，物語，コラムなどは事実と著者の考えや想像が入り交じっているため，光と影が同程度に入っていると考えられる．また，詩，俳句，和歌などは人に様々な情景を想像をさせるため，影となっている部分が多く，光が当たっている部分は少ないと考えられる．このように，文章によって光と影の割合は異なっており，提案システムによって最適な割合となるように文章の作成を支援する．

2．関連研究

これまでにも書かれた文章の読みにくさを評価し，可読性の向上を図るための研究がなされてきた．ディスプレイ上で文章を読みやすくするために，ユーザの興味に近い部分や文章のテーマに近い部分を彩色するシステム [内田 97] や，患者向けの薬の説明書の読みにくさを，漢字の数と 1 つの漢字を構成する文字のつながり数から評価する研究 [酒井 06]，英文の難易度を 1 つの単語中のアルファベットの数や難しい文法の数から判定する研究 [永田 04] があった．これらの研究で

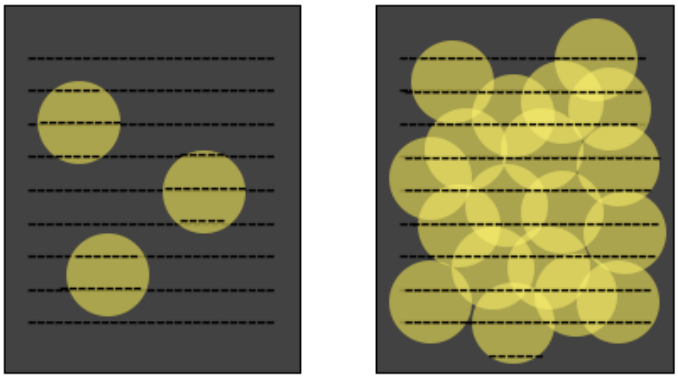

図 1．文章を写真に見立てたときの，光と影の分布イメージ

は文章の見た目の読みにくさを判定しており，意図を伝えるために説明が足りているかとを評価していない．他に，小論文の自動採点システム jerater があり [石岡 02]，このシステムは文の長さや漢字・かなの割合といった文章の修辞，順接と逆説の接続詞の順序の整合，以前に書かれた新聞のコラムとの類似性から論文を採点するものである．この研究では，書かれた小論文の内容と書かれたコラムとの類似性を評価するため，文章の説明が足りているかを評価しない．本研究では文章において説明が足りているかを評価するために，文章のテーマを表す単語とそれ以外の単語の関連を評価し，テーマと関連がある単語が多く使われている文章ほど，説明が足りている分かりやすい文章とする．

3．提案システム構成

提案システムは，ユーザによって作成された文章と，文章のテーマを表す単語集合を入力とする．提案システムは文章のテーマを表す単語集合と，それに関係する文章中の単語集合を用いて，文章の光と影を評価する．システムは文章の光と影の分布状態，光の部分に頻出する単語のリスト，影の部分に頻出する単語のリストを視覚インタフェース上に出力する．

3-1　文章の光と影の評価方法

文章の光と影の評価方法は次の通りとする．文章にはあるテーマが存在し，それに基づき文章が書かれていると考えられる．文章中の各文はテーマと関係があり，テーマを表す単語やそれとの関係が強い単語が多く使われているほど，テーマとの関係が明確に表現されている文と考えることができる．反対に，テーマを表す単語に関係がない単語が使われているほど，テーマとの関係が明確に表現されていない文と考えることができる．したがって，テーマを表す単語やそれとの関係が強い単語を光源とし，光源が及ぶ部分を光，光源が及ばない部分を影と評価する．

3-2　光源となる単語の選択

光源となる単語を選択するために，文章中のそれぞれの単語にテーマとの関連を表すラベルを付与する方法を説明する．本研究は川下りシステム [砂山 06] で提案されている方法を用いて，単語にラベルを付与する．川下りシステムは，あるテーマに関する会話の始まりから終りまでを川下りに例え，会話の流れの中で使われる単語にラベルを付与し，テーマに関する会話の進行状況を把握する支援を行う．あるテーマに関して書かれた文章の可読性を評価することに，川下りシステムを適用可能なため，本研究では川下りシステムを用いて単語のラベル付与を行う．

文章中のそれぞれの単語には表 1 に示す 6 種類のラベルの内，1 つのラベルが付与される．6 種類のラベルは，テーマと関係する単語（TOPIC, FLOW, NEW, TOPIC）とテーマと関係しない単語（BYWAY, FLOOD）の 2 つに分類される．図 2 にそれぞれの単語に付与されるラベルの遷移

表１．各単語に与えられるラベルとその意味

ラベル	意味
TOPIC	文章のテーマとなる単語
FLOW	テーマに関連する既出単語
NEW	テーマに関連する新出単語
INC	テーマに関連しない単語から， テーマに関連する単語として，新たに組み入れられた単語
BYWAY	テーマに関連しない既出単語
FLOOD	テーマに関連しない新出単語

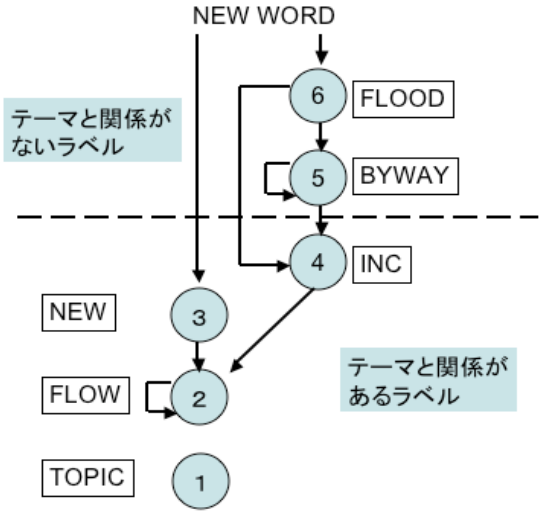

図２．各単語に付与されるラベルの遷移パターン

パターンを示す．テーマを表す単語には TOPIC のラベルが付与され，ラベルが変更されることはない．それ以外の新出単語にはテーマに関連するならば NEW，テーマに関連しないならば FLOOD とどちらかが付与される．新出単語がテーマを表す単語から 10 単語以内に存在していれば，その単語はテーマに関連すると判断する．１回目に NEW のラベルが付与された単語は，２回目に出現したときには FLOW のラベルが付与される．一方，１回目に FLOOD のラベルが付与された単語は，２回目以降，テーマを表す単語に関連すると評価されたときに，一度だけ INC のラベルが付与され，INC のラベルが付与された単語は次の出現時には FLOW のラベルが付与される．

単語に付与されたラベルから，光源となる単語を選択する．本研究ではテーマを表す単語やそれとの関係が強い単語を光源とする．TOPIC, FLOW, NEW, INC のラベルはテーマと関連する単語に付与されるラベルだが，このうち，NEW と INC はテーマと初めて関連する単語に付与されるラベルであり，テーマに関して説明が足りているとは言えない．そこで，TOPIC と FLOW のラベルが付与された単語を光源とする．

3-3　光源の強さの設定

光源には強さがあり，単語によって光源の強さが異なると考えられる．TOPIC はテーマを表す単語であるから，FLOW よりも強い光を放つと考えられる．そこでそれぞれの光源の強さを I_t, I_f とし，ある閾値 T を設け，$T \geqq I\_t > I\_f > 0$ となるよう数値を割り当て，それぞれの光源の強さを表現する．光源の強さはその単語の前後で，照らされる単語の数とする．

3-4　システム出力

光源とその種類を決定した後，文章の光と影の分布状態を評価する．このときにある一定以上の光が当たっている文に頻出する単語のリストと，光が当たっている量が一定以下の文に頻出する単語のリストを作成する．最後に入力された光と影の分布状態と，光が当たっている文に頻出する単語のリスト，光が当たっていない文に頻出する単語のリストを出力する．

3-5　想定するシステムの使用法

表示された光と影の分布状態と２つの単語リストを用い，ユーザは作成した文章を修正する．例えば，ユーザには影となっている部分を修正してもらう．このとき影の部分に頻出する単語を削除するか，影の部分に含まれる文に光の部分に頻出する単語を加えることで，影を減らすことができる．一方，影の部分が減らされ，光の部分が増やされることによって，文章に冗長さが生じると考えられる．そこで，適切な光の量にすべく，文章を修正していく．ここでは，それぞれの文に当たる光の量の上限を設け，その上限を超える文に含まれる単語の内，テーマやテーマとの関係が強い単語を除くことで，光の量を抑えていく．以上の作業によって，テーマとの関係が明確に表現され，かつ冗長でない文章を作成することができる．

４．おわりに

本稿では，文章の可読性を向上する文章作成支援システムを提案した．提案システムでは，文章においてテーマと関連がある単語が多用され，説明がなされた部分を光，テーマと関連がない単語が多用され，説明が不足している部分を影とし，文章の光と影を評価することで可読性を向上する支援を行う．

今後は提案システムの評価実験を行っていく．評価実験では，まず，提案システムが光と影を評価できることを確認する．実験では，被験者によって既存の文章の光と影の部分を評価してもらい，それとシステムの出力結果を比較する．これによって，提案システムが文章の光と影を評価できることを確認する．続いて，提案システムによって，可読性が高い文章を作成する支援が行えることを確認する．実験では，被験者にあるテーマについて，800字程度の文章を書いてもらう．その後，書いてもらった文章を，提案システムを用いず，被験者自身で出きる限り修正してもらう．最後に，被験者に提案システムを用いてもらい，さらに修正すべきところを修正してもらう．提案システムによって，より多くの修正が行われるかを調べると共に，各段階で得られる文章の可読性を被験者に評価してもらう．これによって，提案システムの有効性を確認する．提案システムによって，意図が伝わる文章を作る支援を行い，人間の知的作業の一つである情報発信の効率をあげていきたい．

参考文献

[内田 97] 内田友幸，田中英彦：可読性向上を図る対話的文書自動彩色システム，電子情報通信学会論文誌 D，Vol.J80-D2，No.12，pp.3173-3180，1997.

[酒井06] 酒井由紀子：患者向け説明文書の可読性判定，2006 年度三田図書館・情報学会研究大会，2006.

[永田 04] 永田亮，井口達也，桝井文人，河合敦夫：日本人英語学習者を対象にした英文難易度判定手法，電子情報通信学会論文誌 D，Vol.J87-D2，No.6，pp.1329-1338，2004.

[石岡 02] 石岡恒憲，亀田雅之：コンピュータによる日本語小論文の自動採点システム，電子情報通信学会技術研究報告（思考と言語），Vol.102, No.491，pp.43-48，2002.

[砂山 06] 砂山渡：議論の流れを制御する電子掲示板ー川下りシステムー，第 22 回人工知能学会ことば工学研究会資料，pp.63-70, 2006.