

# ユーザの癖を考慮したネチケットに基づく電子メール作成支援システム

古舘 弥之 榎松理樹 藤田ハミド  
岩手県立大学 ソフトウェア情報学部

E-mail {g031z151, kure, issam}@soft.iwate-pu.ac.jp

## 1. はじめに

近年、電子メールは、利用環境が整備されてきたことやその利便性から、重要なコミュニケーション手段となってきている。E ジャパン協議会が2000年に家庭や個人に対して行なった調査[1]においては、PCを所有している約1200のモニター世帯の内、約70%、携帯電話などを所有している約1200人の内、約33%が電子メールを利用していると回答している。また、同協議会の調査によれば、PCを持っている世帯での電子メールの利用世帯数は、29.3%(97年)、41.7%(98年)、54.5%(99年)、70.2%(00年)と年々増加している。これらの回答に見られるように電子メールは重要なコミュニケーション手段となってきている。

電子メールの利用が増えるに伴い、電子メールに関するトラブルも増加している。財団法人インターネット協会が2002年に電子メール利用経験者約1,700人を対象に行なった調査[2]では、メールが原因でトラブルになった経験がある人は15%弱であり、その主な原因は、「言葉の行き違い」(38%)、「メールの不達(宛先ミスを含む)」(18%)、「メールの遅延」(14%)、「ウイルス」(13%)、「マナー違反」(7%)であった。

このようなトラブルに対し、電子メールを含めインターネットを利用する際のルールやマナーがネットワーク上のエチケット、ネチケットとしてまとめられ、数多く提示されている[3][4]。それらネチケットの内容は、大きく情報を発信する立場と受信する立場にわけることができる。ユーザがより注意すべきであるのは、情報を発信する立場でのネチケットである。メールに関していえば、メールを作成する際に発信する側のネチケットを理解し、気をつけることができれば、トラブルを軽減することが期待できる。しかしネチケットの中には、添付ファイルの容量や機種依存文字といったコンピュータ自体に起因する点も多くあり、ユーザ、特に初心者には理解するのは困難である。またユーザの経験や良識に頼る部分も多く、ユーザへの負担が大きい。

このような負担を軽減するために利用できる技術としては、自然言語処理分野で研究されている文書校正支援があげられる。文書校正支援としては、キ

ャノンシステムソリューションズの文章チェックライブラリ「K-Tech Proofread」[5]や、Just Systemの文書構成支援ツール「Just Right!/R.2」[6]、言語理解研究所の文書校正エンジン[7]などがあるが、これらは文章が対象であり、電子メールそのものには特化していない。そのため、より円滑な電子メールでのコミュニケーションの実現のため、ネチケットを意識した電子メール作成支援が求められている。

以上のような背景から、本研究では、電子メール作成に焦点を絞り、ネチケットに基づく文書作成支援を行なうシステムを提案する。本システムは、ユーザが作成する電子メールに対して、ネチケットの観点から問題点とその理由を提示し、修正を促す。これにより、適切なメールの作成を支援するとともに、ユーザのネチケットへの意識を高めることを目的とする。また、ネチケットの理解や利用には個人差が大きいことから、ユーザの作業履歴から見落としやすい項目を癖として発見し、それを反映することで支援効率の向上を図る。なお現時点では、本システムはユーザに指摘を行なうことを目的とし、自動訂正は行なっていない。

## 2. 電子メール作成支援システム

### 2.1 システムの構成

図1に本システムの構成図を示す。本システムは、作成支援部と癖発見部から構成される。作成支援部では、ネチケットルールベース(RB)のルールとユー

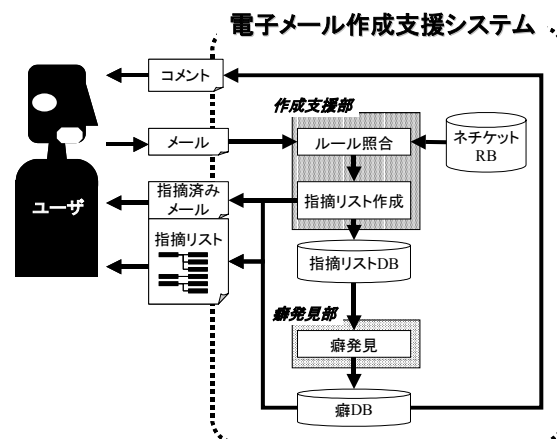


図1: システムの構成図

ザの入力したメールとを照合し、指摘箇所と指摘理由（照合されたネチケットルール）から構成される指摘リストを作成する。このリストと指摘箇所を追加した元のメールとをユーザに提示し、メールの修正を促すことで文書作成支援を試みる。さらに癖発見部においてユーザごとに保存した指摘リストから出現頻度に基づいてユーザの癖の発見を試みる。ここで癖とは、「ユーザが見落としやすい、軽視しやすい、または意識していないネチケットの項目や項目間の関係」のことを指す。この癖を文書支援部で利用することにより、ユーザに合わせた支援を行なうことができ、支援効率の向上が期待できる。

## 2. 2 ネチケットルールベース

ネチケット RB には、チェックを行なうネチケットの項目がルール形式で記載されている。各ルールの条件部には、ネチケットの判定の基準となる文書の特徴が記載されており、結論部にはコメント作成用のテンプレートが与えられている。

RB に蓄えられるネチケットは、文献[3][4]のように提示されているもののなかから、メールの新規作成時における表層情報(表現)に関連する点のみを対象とする。理由としては、メールの新規作成がメールによるコミュニケーションの起点となること、ユーザ自身が関与できる点が大きいこと、「言葉の行き違い」がトラブルの大きな原因としてあげられていること、コンピュータでの処理がしやすいなどがあげられる。

## 2. 3 作成支援部

作成支援部では、ユーザが作成したメールとネチケット RB 内のルールとの照合を行なう。処理の流れを次に示す。

- ① ユーザは、システムを用いてメールを作成する。この時、「指摘すべきユーザの癖」がある場合、それをコメントとしてユーザに提示する。
- ② 入力されたメールに対し、文字単位、形態素単位での解析を行なう。文字単位での解析では、メールの文字数、一行単位の文字数、行数、文字の種類といった文字単位で得られる情報を抽出する。形態素単位での解析では通常形態素解析を行ない、文書を形態素に分割する。
- ③ ②で抽出した情報とネチケット RB 内の各ルールの条件部とを照合し、指摘箇所を発見する。照合は、全ルールに対し行なう。照合した指摘箇所はその理由となるルールとを組み合わせ、指摘リストに追加する。このようにして作成した指摘リストを指摘箇所の情報を付加したメール文書とともにユーザに提示する。

指摘リストについては、ユーザの利便性を考え、下位ノードがより具体的な項目を示す形の木構造リストで表示する。さらに指摘理由や後述する発見した癖を提示することでユーザへのネチケットの理解を深めることを試みる。

また作成した指摘リストは、個人ごとに用意する指摘リスト DB に格納する。

- ④ ユーザがメール作成を続ける場合は①へ戻る。それ以外は終了する。

以上の処理を通して、メールの作成支援を試みる。

## 2. 4 癖発見部

癖発見部は、作成支援部の処理が終了後、処理を実行する。癖発見部では個人ごとに用意した指摘リスト DB 中の指摘リストから、出現頻度をもとに次に示す癖の抽出を行なう。これらの癖は、ユーザごとに用意する癖 DB に保存しておき、支援内容に合わせて利用する。

- ① 見落としがちな項目：指摘リスト DB 内にある全指摘リストの指摘理由に対し、一定の回数以上に出現するものをユーザが見落としがちな項目と判断し、癖として抽出する。これをメール作成時にコメントとして提示することで、ユーザのネチケットへの意識の向上を図る。
- ② 軽視しやすい項目：指摘リスト DB 内にある全指摘リストの指摘理由を出現回数の多い順に順位付けを行なう。この順位を軽視しやすいと判断し、癖として抽出する。修正リストをユーザに提示する際に、この順位で項目を並びかえることによりユーザが軽視しやすい項目への意識を高める。なお初期状態は、初期設定に従う。
- ③ 関連する項目：修正リストの内容を時系列でとらえ、関連する項目の発見を試みる。ここで関連する項目としては、(1)一緒に出現することが多い、(2)一方が解消されるともう一方も解消される、(3)一方が解消されると逆に出現する、という3種類の関係である。各項目間において、それぞれの関係が一定回数以上おきた場合、それら項目間の関係を、ユーザが意識しにくい関連項目と判断し、癖として抽出する。その癖を修正リストの中にコメントとして提示することにより、局所的ではなく大局的な視点を促す。またこれにより、一回の作業において、より多くの内容を修正することで、修正回数の軽減を図る。

以上の処理で癖の発見を行ない、作成支援部で利用することで、支援向上を図る。

### 3. 評価実験

#### 3.1 実験概要

本システムの有効性を評価するために、プロトタイプを用いた次の二つの実験を行なった。

**実験 1:** 本実験では、作成したプロトタイプの指摘を評価者が評価し、システムの出力の妥当性を検証する。妥当性が高いほどシステムが有効であると判断する。

**実験 2:** 本実験では、同じデータに対して評価者が行なった判断とシステムの出力との比較を行ない、どの程度一致するかを評価する。なお、評価者の判断結果とシステムの出力結果とが一致したかの判定は、評価者が行なう。また評価者に対しては、事前にシステムに関する情報は与えておらず、その個人の基準で判断している。

これらの実験の結果、人の判断結果とシステムの出力の一致数が多ければ多いほど、本システムの有効性が高いと評価する。

実験で利用したプロトタイプは Windows2000 上で、JAVA Version1.4.2 を用いて開発している。形態素解析部については、京都大学大学院の黒橋らが開発した日本語形態素解析システム「JUMAN」[8]を利用している。またネチケット RB には、文献[3][4]などで提示されているネチケットから、前述の条件を満たすものを中心に抽出、整理したものを与える。その概要を表 1 に示す。表 1 で示すように今回はチェック対象を大きく 5 つに分類し、各対象に対し具体的な項目を設定している。

表 1: ネチケット RB の構成

チェック対象	具体的な項目
表層的な文章表現	漢字比率、句読点の統一、語尾の統一、接続詞の重複、漢字の種類、機種依存文字の制限、括弧の対応
不快感	差別語、放送禁止用語、命令語
レイアウト	文章全体の長さ(長文)、段落の長さ、改行までの長さ、一文の長さ
文書内情報	公開情報、署名の構成、件名の長さ
その他	送信者と受信者の明記

実験データとしては、観光スポットの紹介という同一目的で書かれた電子メール 53 件分を用いた。本データは 18~20 歳の男性 4 名、女性 49 名が作成している。また作成者は事前にネチケットの説明をうけているが、システムについては知らされていない。実験 1、実験 2 とも本データを用いている。

なお、今回の実験においては、一人のユーザが複数回利用しているものではないため、癖の発見については実施していない。

#### 3.2 実験結果

実験 1 で行なった、システムの出力結果に対する評価者の評価結果を表 2 に示す。表 2 において「適切」とは、システムが出力した指摘個所のうち、評価者が適切と判断した指摘、「不適切」は不適切と判断した指摘の数である。「不足」は評価者が、表層情報に関して追加した指摘の数であり、指摘総数はこれらの合計を示す。指摘総数に対する比率は各評価の数を、指摘総数で割って求めた値である。

表 2: (実験 1) システムの指摘の評価

評価	数	指摘総数に対する比率
適切	76	74.5%
不適切	20	19.6%
不足	6	5.9%
指摘総数	102	

表 3 に、実験 2 で行なった、評価者の判断とシステムの出力との比較結果を示す。表 3 において、「指摘した」とはそれぞれが提示した内容をさし、「指摘せず」は指摘しなかった内容を示す。

表 3: (実験 2) 評価者の評価との比較

		評価者の判断	
		指摘した	指摘せず
システムの出力	指摘した	22	74
	指摘せず	41	0

#### 3.3 評価・考察

実験 1 の結果、システムの出力のうち、評価者が適切と判断した指摘個所は 76(約 74.5%)と高い割合を占めた。指摘内容としては、文字数や行数、一文の文字数、機種依存文字など比較的簡単な部分であった。しかし簡単な部分、言わば人が見落としやすい部分を指摘できたと考えられる点から本システムの効果はあると考える。

一方、約 20%が不適切な指摘であった。不適切と判断された個所としては、機種依存文字の検出間違い、署名の認識間違いなどが挙げられる。機種依存文字については単純に設定の問題であったが、署名については、メール作成者がネチケット RB 構築時には想定していなかった利用方法を行なったため、正しく認識ができなかったことが原因であった。また不足とされた部分は、実験に利用したメールの文章が短かったために、開発時に想定した各種ルールの設定と適合しなかった点が大きな原因であった。これらの点から、実際のメールにおける表現は多様であり、作成者、特に初心者であるほど、システム

構築時の想定から逸脱したメールを作成しやすく、それらの判定が難しいことが分かった。そのため、作成者のレベルや状況、メールの大きさに合わせたネチケットのレベルわけや現在のルールとは異なるネチケットのモデル化の検討が必要である。

実験 2 の結果からシステムの性能を評価すると、再現率は 34.9%、適合率は 22.9%、正解率は 16.1%、カバー率は 46%であった。各値とも高い値にはなっていない。

評価者が指摘し、システムが指摘できなかった多くの部分は、メールの目的が明確になっているか、送信者と受信者との社会的関係を考慮した表現となっているか、といった現在利用している表層情報のみでは判断が困難な点であった。これらの点に対応するためには、構文解析結果や分かち書き、テンプレート、メールの持つ文書情報といった現在利用していない情報を利用するなどの検討が必要である。

一方、システムの指摘したが評価者が指摘しなかった点のなかには評価者が見逃した点がいくつか含まれている。現在精査中であるが、それらは比較的簡単な部分が多い。このことから、人が見逃しがちな簡単な点の指摘を本システムが行なえると考えることができる。

### 3. 4 今後の課題

今回の実験結果から、開発時の目標としている表層レベルでの指摘が行なえる可能性が示せた。また比較的簡単な表層部分の指摘は、実験 2 の結果からも伺えるように、人が見落としやすい点であると考えられ、その点での支援の可能性、有効性があると考えられる。

一方、十分な指摘ができなかった点は、現在利用している表層情報のみでは難しい点やシステム開発時の想定外の表現に起因するものであった。これらに対応するためには、実際のメールを解析し、その中から経験則を発見することや、ユーザのレベルにあわせたネチケットのレベルわけなどが考えられる。また、メールのもつ文書情報やメールの目的にあわせたテンプレートの導入、社会的関係の考慮、現在行っていない構文解析などの利用などを検討する必要がある。

さらに今回実験できなかった癖発見および癖の利用についても実験し、その有効性を評価する必要がある。現在の癖の発見は出現数による単純なものであるが、指摘箇所に関するより詳細な情報を利用することや、ルール発見を行なう範囲を変更し、ルール生成や発見技術を援用することで、癖のみではなく、ネチケット RB に反映できる経験則の発見も期待できる。

### 4. おわりに

近年、電子メールはユーザの増加とともに重要なコミュニケーション手段の一つとなっている。それとともに電子メールに関するトラブルも多くなっており、その中には文書表現に起因するものも多い。現在、ネチケットとして電子メールに関するマナーが提示されているが、コンピュータ自体に起因する点も多く、ユーザ、特に初心者にとっては、それを反映した電子メールを作成することは難しい。

本稿では、このような問題に対し、電子メール作成時にネチケットに基づく支援を行なうシステムを提案した。プロトタイプを用いた評価実験の結果、指摘の約 75%が妥当であると判断され、本システムの有効性が示せた。また同じメールに対する人の判定結果との比較については、再現率 35%、適合率 23%と低い値であったが、人が見落としした比較的簡単な箇所の指摘を行なえたことから、人が見落としがちな単純なレベルでの支援として有効に働く可能性を示せた。

今後の課題としては、実際のデータの解析、今回指摘できなかった部分に対する対策方法の検討、癖発見部の作成と利用を行ない、検証を通してシステムの改善を図ることが挙げられる。

### 参考文献

- [1] E ジャパン協議会 : <http://www.ejf.gr.jp/>
- [2] 財団法人インターネット協会 : <http://www.iajapan.org/>
- [3] ネチケットガイドライン(RFC1855)日本語版 2.1.1 電子メールのガイドライン : [http://www.cgh.ed.jp/netiquette/rfc1855j.html#2\\_1\\_1](http://www.cgh.ed.jp/netiquette/rfc1855j.html#2_1_1)
- [4] バーニジア・シャー著 松本功訳：ネチケット—ネットワークのエチケット—、ひつじ書房(1996)
- [5] 文章チェックライブラリ「K-Tech Proofread」<http://ps.canon-sol.jp/sv/product/proofread.html>
- [6] 文章校正支援システム「Just Right!/R.2」 : <http://www.justsystem.co.jp/justright/>
- [7] 文書校正エンジン : <http://www.ilu.co.jp/>
- [8] 日本語形態素解析システム「JUMAN」 : <http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman.html>