[DRAFT]

[DRAFT]

本記事は，情報処理学会誌「情報処理」（Vol.53, No.3, pp.202–203, 2012）に掲載されているものの DRAFT 版です．

編集にあたって

乾 孝司†

今月は，人間が書いたり，話したりする言葉を計算機で処理する自然言語処理技術に関する特集です．自然言語処理は，従来，新聞記事に代表されるような整った言語データを主に研究や実処理の対象として選んでいましたが，近年では，インターネット上にあふれる，整っているとは言い難いさまざまな言語データに対する処理が要請されるなど，時代と共にそのニーズは枝分かれしています．

このような背景から，本特集では，従来の自然言語処理では，あまり主役として扱われなかった言語現象，整った言語現象と比べてみると，どこか不自然さを感じられる言語現象およびその周辺の話題を「不自然言語処理」と名付け，スポットを当てました．

そもそも「自然言語処理って何だかわからない」という読者の皆様にも楽しんで頂けるよう，できる限り身近で，馴染み深い不自然さをもつ言語現象をテーマに選んだつもりです．本編の読後に，「あーこういうものも自然言語処理なんだ」と，少しでも肌で感じて下されば幸いです．

本特集は，７つの話題から構成されていますが，それらは大きく３つのパートに分かれています．はじめのパート（２件）は，情報通信技術の発展と共に生まれた不自然な現象に関する話題です．

まず，「1. 顔文字処理（Michal Ptaszynski)」では，メール等のテキストベース・コミュニケーションの場において使われる顔文字に関する処理について解説して頂きます．顔文字がどのような目的で，どのように誕生したのかについて，また，顔文字処理の最新技術として，著者が開発中の CAO システムを中心に紹介して頂きます．

次に，例えば英語では，”I love you”のように，単語間の境界情報が明示されますが，日本語はそうではありません．そのため，日本語の解析処理では，通常，まず始めに，文章を単語に分割する処理をおこないます．しかし，先程の顔文字も含め，近年では，ソーシャルメディアの普及と共に，これまでは存在しなかった語や言語的表現が日々生産されており，単語分割処理を困難にしています．「2. 新しい語・崩れた表記の処理（笹野遼平，鍛治伸裕)」では，このような問題に対処する方法として，著者らの最新の研究を解説して頂きます．

続いてのパート（３件）は，母語あるいは第２言語の修得段階で生じる不自然さに関連する話題です．これらの話題の背景には，社会の国際化の他に，以前までは困難だった言語学習過程における言語データの収集や分析が，情報インフラの整備と共に効率よく実現できるようになった，社会の情報化という側面があるようです．

まず，「3. なんで日本語はこんなに難しいの？(水本智也，小町守)」では，日本語学習者向け支援システム，およびそれらを構築するために必要な言語データを取り巻く状況について概観した後，学習者が書いた誤りを含むテキストの自動解析の難しさ，さらには，テキスト中の誤りを自動訂正する最新の手法・システムを紹介して頂きます．

続く「4. 英語学習支援（乙武北斗)」は，先程とは言語の方向が違って，日本人のための英語学習支援に関する解説です．日本語と英語がもつ文法の違いに着目し，英語学習において日本人学習者が誤りやすい文法項目を整理したうえで，それら項目の誤りをテキストから自動検出し，訂正する手法について，過去から現在までの手法の変遷を中心に概説していただきます．

† 筑波大学 システム情報系

さて，先の２件は，母語を習得している人が，母語以外の言語を第２言語として学習する状況でしたが，続く「5. 日本語学習児の初期語彙発達（小林哲生，永田昌明）」は母語習得に関する内容です．一般には１才を迎えるころから，徐々に言葉を覚え始めるようですが，１才前後の幼ない子供たちの言語データを収集することは，先の第２言語学習時の事案とはまったく質の異なる困難さがあるようです．本稿では，著者らの最近の活動を中心に，言語データ自体に関する研究，および幼児の語の理解に関する実験をご紹介いただきます．

最後のパート（２件）は，本特集でいう不自然言語処理の応用事例に関する話題です．どちらも，情報通信技術の発展に付随する形で生まれた不自然な言語現象を多く含むマイクロブログを処理対象としています．

マイクロブログは，情報発信の手軽なツールとして，近年多くのユーザを集めています．「6. Twitter からの情報抽出（荒牧英治，橋本泰一）」では，マイクロブログの代表サービスの一つである Twitter から，感染症や被災文化財等の特定の情報を見つけ出す応用事例を紹介して頂きます．特に，マイクロブログの即時性に着目することで，いつ・どこで・何が起きたかが，わかりやすく整理された形で提示される様子は大変興味深いです．

最後の記事「7. ANPI_NLP（村上浩司，萩原正人，Graham Neubig，松林優一郎）」は，昨年，甚大な被害をもたらした東日本大震災において立ち上がった，震災時安否情報確認支援プロジェクト「ANPI_NLP」について，プロジェクトで主導的な立場にあった著者らに解説を依頼したものです．災害時，中央集権的でないソーシャルメディアは，頑健な情報通信基盤を提供した反面，いま必要な情報を，その情報が必要な人へどのように伝えるかが重要な問題でした．言語処理技術を用いてこの問題に立ち向かい，情報の交通整理に奔走したプロジェクト・メンバらの活動を，技術的側面も交えながら，ご紹介いただきます．

それでは，「不自然言語処理」の世界をどうぞお楽しみ下さい．