

中国で活動する企業のための社名翻字システムに関する研究

中村健二[†] 田中成典[†] 増満光[‡] 西江将男[‡] 二宮倫子[†]

関西大学総合情報学部[†] 関西大学大学院総合情報学研究科[‡]

1. はじめに

近年、中国市場は世界から注目を集めている。我が国の企業も、東証一部上場企業の 6 割以上がすでに中国において活動しており、今後さらに多くの企業が中国市場に進出することが予想される[1]。漢字名を持たない企業が、新たに中国市場に進出する際、中国語の漢字名を用意する必要がある。しかし、漢字は表意文字であるため、意味と音の双方を考慮する必要がある。中国語未習者による中国語翻字は困難である。そのため、専門家に依頼して名付けるのが一般的であるが、コストが必要となる。また、既存研究[2]-[5]では、音訳のみに特化した翻字の手法を提案しているが、意味訳、そして意味訳と音訳の混ざった翻字には対応できていない。さらに、業種を考慮しておらず、自動車会社では“速い”，食品会社では“美味しい”といった業種別の相応しい漢字が選定できていない。そのため、本研究では、意味訳、音訳、そして意味訳と音訳の混ざった翻字に対応し、さらに業種を考慮することで、社名を中国語翻字する手法を提案する。

2. 研究の概要

本研究では、社名を中国語翻字する手法を提案する。本システム (図 1) は、1) 意味と音を考慮した翻字のための辞書構築機能、2) 意味を考慮した翻字機能、3) 音を考慮した翻字機能で構成される。入力データは、翻字対象の社名とし、出力データは、中国語社名の候補群とする。

2. 1 意味と音を考慮した翻字のための辞書構築機能

本機能では、意味を考慮した翻字に使用する日中対応語 DB、音を考慮した翻字に使用するローマ字・ピンイン対応関係 DB を構築する。日

Research on Transliteration from Japanese Company Names into Chinese Names

[†] Kenji Nakamura, Shigenori Tanaka, Noriko Ninomiya
Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

[‡] Hikaru Masumitsu, Masao Nishie
Graduate School of Informatics, Kansai University, 2-1-1 Ryouzenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

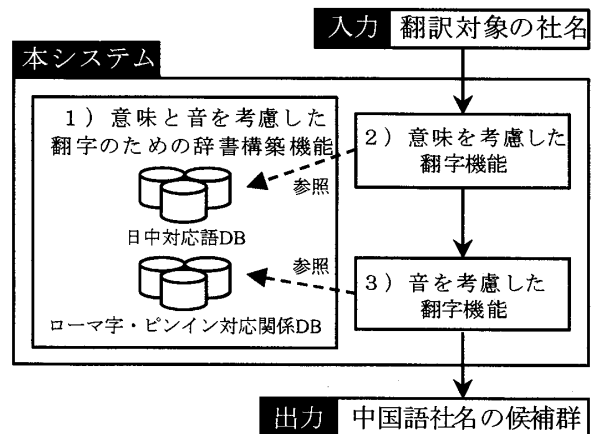


図 1 システムの概要

中対応語 DB の構築では、まず、あらかじめ業種別に収集した中国語社名を形態素に分割する。次に、業種ごとに中国語形態素の TF 値を、社名全体において IDF 値をそれぞれ算出し、各業種の特徴となる中国語形態素を抽出する。そして、抽出された各業種の特徴となる中国語形態素を中日辞典により日本語に変換する。最後に、中国語形態素、中国語形態素に対応する日本語形態素と TF-IDF 値を辞書 DB に登録する。ローマ字・ピンイン対応関係 DB の構築では、まず、日中英社名・ブランド名辞典から業種別にカタカナの社名を取得する。そして、カタカナの社名をローマ字に変換し、ローマ字、ローマ字と対応したピンイン、漢字と出現数を辞書 DB に登録する。

2. 2 意味を考慮した翻字機能

本機能では、日中対応語 DB を参照し、入力した社名に対して意味を考慮した翻字を行う。まず、入力した社名を形態素に分割する。そして、形態素が地名であれば、日中辞典を用いて中国語の地名に変換する。地名でなければ、日中対応語 DB を参照し、形態素と対応する中国語を TF-IDF 値の高い順に選定し、中国語に翻字する。

2. 3 音を考慮した翻字機能

本機能では、意味を考慮した翻字機能において、翻字不可能であった形態素に対して音を考

慮した翻字を行い、既に翻字された結果がある場合、結合して出力する。まず、形態素をローマ字へ変換する。次に、ローマ字・ピンイン対応関係 DB を参照し、ローマ字をピンインへ変換する。そして、企業の業種に合わせて、出現数の多い順にピンインと対応する漢字を選択して翻字する。最後に、意味を考慮した翻字機能の結果と結合し、順位付けして中国語社名の候補群を出力する。

3. システムの実証実験と考察

本システムの有用性を実証するため、中国語社名を持つ日本企業の社名 30 件をテストデータとし、本研究手法を用いた場合の翻字精度を評価した。また、出力する中国語社名候補が適切であるかを検証するため、中国人被験者にアンケート調査を行った。

3. 1 実証実験

実証実験では、本研究で採用している“意味、音と業種を考慮した翻字”の有用性を実証するため、既存研究で採用している“音のみを考慮した翻字”と“意味と音を考慮した翻字”との翻字精度の比較実験を行った。実験方法は、日本語社名 30 件を入力データとして、本システムにて翻字した中国語社名候補上位 50 件以内に実際の中国語社名が含まれるかどうかを確認し、その件数で評価した。

同時に、本システムが出力する中国語社名候補が適切であるかを検証するため、中国人被験者 5 名を対象にアンケート調査を行った。アンケート調査は、10 社の企業に対して、本システムにより翻字した中国語社名上位 30 件を提示し、中国人の感性に準じた社名かどうか 5 段階評価で問う形式とした。さらに、本システムの改善点を把握するため、提示した候補群に対する意見や感想を記述する自由回答欄を設けた。

3. 2 結果と考察

翻字精度の評価実験の結果を表 1 に示す。“意味、音と業種を考慮した翻字”の結果が、最も良いことがわかった。これは、本手法が意味や音を考慮した様々なパターンの中国語翻字に対応できたためであると考えられる。

アンケート調査の結果を表 2 に示す。表 2 より、上位であるほど平均評価が高い結果となり、本システムの有用性が高いことを実証した。しかし、自由回答欄では“候補群の社名に斬新さを感じられない”という意見が複数みられた。これは、本手法が既存の社名データに依存しているためであると考えられる。また、企業理念や、社会的イメージなど企業の印象を反映した翻字が行えていない問題もある。

表 1 翻字精度の評価実験

	上位 50 位に含まれた件数
音のみを考慮	11
意味と音を考慮	20
意味、音と業種を考慮	23

表 2 アンケート調査の結果

	平均評価
上位 30 位以内	2.7
上位 20 位以内	3.5
上位 10 位以内	3.9

4. おわりに

本研究では、意味訳、音訳、そして意味訳と音訳が混ざった翻字に対応し、業種の特徴を考慮した漢字を選定する社名の中国語翻字手法を提案した。実証実験により、本研究の有用性を実証した。しかし、本研究では、既存の社名に依存する手法を採用しているため、斬新な社名を出力できない問題や企業の印象を十分に考慮した翻字を行うことができない問題がある。そこで、今後は中国人の感性に準じた社名の特徴を調査、分析することや企業の印象が含まれる適切な情報源を取得することでこれらの問題を改善し、さらなる精度の向上を目指す。

参考文献

- [1] 21 世紀中国総研：中国進出企業一覧 上場企業篇 <2009-2010 年版>、蒼蒼社、2009.4.
- [2] 黄海湘、藤井敦：中国語への翻字における関連語抽出の効果、自然言語処理研究会報告、情報処理学会、Vol.2007, No.7, pp.9-16, 2007.1.
- [3] Goh, L., Asahara, M. and Matsumoto, Y. : Chinese Unknown Word Identification Based on Morphological Analysis and Chunking, The Special Interest Group Notes of Information Processing Society of Japan, Information Processing Society of Japan, Vol.13, pp.7-12, 2003.5.
- [4] Li, H., Zhang, M. and Su, J. : A Joint Source-Channel Model for Machine Transliteration, Proceeding of Association for Computational Linguistics 2004, Association for Computational Linguistics, Vol.2004, pp.160-167, 2004.7.
- [5] シューリーリー、藤井敦、石川徹也：言葉の印象に基づく自動翻字手法、自然言語処理研究会報告、情報処理学会、Vol.2006, No.1, pp.103-110, 2006.1.