

英語学習支援ロボットののためのニックネーム自動生成

若井 祐樹[†] 永田 亮[†] 船越孝太郎^{††} 中野 幹生^{††}[†] 甲南大学 ^{††} (株) ホンダ・リサーチ・インスティテュート・ジャパンE-mail: [†]{si871136,rnagata}@center.konan-u.ac.jp., ^{††}{funakoshi,nakano}@jp.honda-ri.com

1. はじめに

2011年度から小学校英語の必修化が始まり、小学校における英語活動の重要性が高まっている。一方で、小学校教員は、英語教育法を専門的に学ぶ機会が少なく、指導に不安があるという意見も多い[3]。外国語指導助手(ALT)による授業補助を利用することも考えられるが、質の高いALTを数多く確保するのは困難である。今後、安定した小学校英語を継続していくためには、小学校教員だけで授業が行える方法を確立していくことが好ましいといえる。

この課題の解決策としてロボットの利用が提案されている。すなわち、ロボットによる英語学習支援や英語指導支援である。例えば、Leeら[5]や寄田ら[6]は、英語でコミュニケーションを行う英語学習支援ロボットを提案している。また、Changら[1]は、ロボットの特性と学習効果の関係を調査している。現状では、技術的な課題が残されている部分もあるが、ロボットによる支援の有効性が徐々に示されつつある。

ロボットを利用した英語学習支援における重要な課題の一つとして、児童とロボットの間の円滑なコミュニケーションが挙げられる。コミュニケーションの第一歩として、ロボットは児童の注意を引きつけられなければならない。そのためには、違和感なく児童に話しかけられることが望ましい。また、児童の興味を持続させることも学習活動において重要である。興味を持続させることができなければ、学習支援や指導支援そのものが成立しないことは容易に想像できる。

このような背景を受け、本研究では、ロボットと児童の円滑なコミュニケーションを促進する手段としてニックネームに着目する。具体的には、児童の(日本語の)名前から、英語のニックネームを生成し、児童とロボットのコミュニケーションに利用することを目指す。例えば、児童が自分の名前をロボットに伝え、ロボットが英語のニックネームで児童に呼びかけるといったように応用が可能である。また、呼びかけだけでなく、名前を基にした短い会話によるコミュニケーションなどにも応用可能である。一例として、次のような会話を挙げることができる：

児童: 僕の名前は、利之です。ロボット: Hi, Toshiyuki. Can I call you Tossie?

児童: Yes!

ロボット: Tossie, do you like cookie?

この例では、児童の名前“利之”から、ロボットは、ニックネーム“Tossie”を生成している。また、“Tossie”と韻を踏む単語“cookie”を使い、会話を続けている。この例のように、ニックネームを介することにより、児童はロボットに親しみがわきやすいと期待できる。実際、神田ら[4]は、名前の呼びかけが児童の興味を引くことに効果的であることを報告している。ただし、本研究で対象とするのは英語学習支援であるため、日本語の名前ではなく英語風のニックネームで呼びかけることが望ましい。更に、英語のニックネームを利用した英語学習方法も提案されており、ニックネームの利用範囲は広い。例えば、教材[2]では、児童のニックネームを織り交ぜた英文を読み上げる英語活動が紹介されている。

そこで、本稿では、日本語の名前から、英語のニックネームを生成する方法を提案する。具体的には、規則に基づく手法、辞書検索に基づく手法、両手法を統合した手法の3種類を提案する。規則に基づく手法では、英語のニックネームの分析から得られた規則により日本語の名前をニックネームに変換することで生成を行う。辞書検索に基づく手法では、日本語の名前に似た英語の名前を辞書から検索し、その英語の名前とペアとなるニックネームを生成結果とする。

2. 提案手法

2.1 前処理

入力となる日本人の名前は、漢字やひらがなから成るため、そのままでは英語のニックネームに変換するのは困難である。そこで、前処理として、日本人の名前をローマ字表記に変換する。本研究では、ローマ字変換にKAKASI^(注1)を用いる。なお、以下では、説明を明確にするため、ローマ字へ変換した日本人の名前を日本名と表記することにする。また、特に断わらない限り、英語の名前、生成されたニックネームをそれぞれ英語名、ニックネームと表記する。更に、母音と子音をそれぞれVとCで表記する。ただし、本研究では、Cには、“y”と“w”を含まないとする。

もう一つの前処理として、日本名の表記のバリエーションを吸収するための正規化を行う。例えば、日本語の“つ”の音は、“tsu”とも“tu”とも表記されるが、これを“tu”に統一する。また、辞書検索に基づく手法では、英語名と英語ニックネームの辞書を利用するが、この辞書に対しても同様な正

(注1): <http://kakasi.namazu.org/>

規格を行う。正規化に用いる規則は付録に示す。

2.2 規則に基づく手法

規則に基づく手法では、英語ニックネームの分析から得られた規則により、日本名を変換することでニックネームを生成する。英語ニックネームのデータ^(注2)879件を分析したところ、ニックネームは、

基本規則：ベース + 境界文字 + 接尾辞

の形をとることが多いことが判明した。例えば、英語名“Adelle”では、

Ad (ベース) + d (境界文字) + ie (接尾辞)

となり、対応するニックネームは“Addie”となる。

そこで、本手法でも、この基本規則をニックネーム生成に利用する。ベースは、日本名の部分文字列から成り、ニックネーム生成の基礎となるものである。境界文字とは、日本名において、ベースとベース以降の文字列との境界にあたる文字のことを示す。より厳密には、ベースの最後の文字とする。また、接尾辞は、ニックネームの語尾に付加されやすい文字列のことである。本研究では、代表的な接尾辞“ie”を用いることにする。

ベースと境界文字の決定方法は次の通りである。日本名の先頭文字から最初のVCまでを取り出す。このとき、先頭からVCまでをベースとする。また、Cを境界文字とする。例えば、“Tosiyuki”の場合、下線部が最初のVCであることから、ベースが“Tos”，境界文字が“s”となる。ただし、ベース内に、母音連続(VVC)が存在する場合は、後の母音を削除する。これは、ニックネームでは、母音の連続を避け、簡潔に発音できるようにする傾向が見られるためである（例えば、英語名“Aileen”では、“Ai”を“A”に短縮し、ニックネームを“Allie”にする）。

ベースと境界文字が決定されると、上述の基本規則を用いて、ニックネームが生成できる。“Tosiyuki”に対しては、

Tos (ベース) + s (境界文字) + ie (接尾辞)

により、ニックネーム“Tossie”を生成する。

以上が規則に基づく手法の基本であるが、境界文字が“n”の場合は、次のような例外処理も行う。なぜなら、“n”は、後続する文字によりナ行の発音にも、“ん”の発音にもなるため、特別な扱いが必要であるからである。ナ行の場合は、上述の基本規則で対応する。一方、“ん”の発音となるとき、すなわち、“n”の次の文字がCであるときに例外処理を行う。このとき、“n”の次の文字を隣接文字とすると、

例外規則：ベース + 隣接文字 + 接尾辞

で、ニックネームを生成する。例えば、“Rintarou”の場合、

Rin (ベース) + t (隣接文字) + ie (接尾辞)

より、“Rintie”を生成する。もし、この例外規則を使用せずに、基本規則を使用すると、“Rinnie”が生成され、日本名が想像しにくいニックネームになってしまう。

後処理として、ニックネームの整形を行う。なぜなら、生成されたニックネームに英語綴りとしては希な文字列が含まれる場合があるからである。例えば、“Mihhie”における“hh”は、英語の綴りでは希であるので、“ff”に変換して、“Miffie”を得る。ニックネーム整形用の規則も付録に示す。

2.3 辞書検索に基づく手法

辞書検索に基づく手法では、ニックネーム生成問題を検索問題として解く。準備として、英語名と英語ニックネームがペアになった辞書を用意する。このような辞書は、Webから容易に取得可能である^(注3)。この辞書から、日本名に類似した英語名を検索し、対応するニックネームを生成結果とするというのが基本的な考え方である。辞書検索に基づく手法を利用することで、基本規則に基づかないニックネームを生成することが可能になり、生成結果にバリエーションを持たせることができる。例えば、“Mitsuki”に対して、辞書検索に基づく手法では、“Mitch”を生成できる（規則に基づく手法では“Mittie”が生成される）。また、規則に基づく手法では、ベースが取得できず、ニックネームの生成できない場合がある。そのような場合に、規則に基づく手法を補う役目も果たす。例えば、“Kou”のようにVCがない日本名に対しては、規則に基づく手法では、ニックネームの生成ができないが、辞書検索に基づく手法では“Coutie”などが生成できる。

辞書検索に基づく手法では、前処理として、辞書中の英語名の正規化を行う。具体的には、日本名では使われない英語名のアルファベットを対応する日本名のアルファベットに変換する。例えば、英語名では、サ行の音を“c”を用いて表すことがあるが、日本名では通常“s”を用いるため、“c”を“s”に変換する。同様に、ラ行の“l”を“r”に変換する（変換規則の詳細は付録に示す）。この変換により、より正確に日本名で英語名を検索することが可能となる。

上述の通り、辞書検索に基づく手法では、日本名に類似した英語名と英語ニックネームを辞書検索する。そのためには、日本名と英語名との類似度および日本名と英語ニックネームとの類似度を定義する必要がある。日本名と英語名との類似度は、日本名と英語名との先頭からのマッチ文字数と定義する。例えば、日本名“Mituki”と英語名“Mitchell”では類似度3となる。同様に、日本名と英語ニックネームとの類

(注2) : <http://usefulenglish.ru/vocabulary/>の“Men's Names and Nicknames”と“Women's Names and Nicknames”を用いた。

(注3) : 例えば、<http://www.censusdiggins.com/nicknames.htm>などが利用可能である。

似度は、日本名と英語ニックネームとの先頭からのマッチ文字数と定義する。例えば、“Mituki”と英語のニックネーム“Mitch”では類似度3となる。

これらの類似度に基づき辞書検索を行い、ニックネームの生成を行う。まず、日本名に対して、類似度が最大となる英語名を辞書検索する。検索された英語名とペアになったニックネームを生成候補とする。次に、生成候補であるニックネームと日本名との類似度を求め、最大となるニックネームを生成結果として出力する。類似度が最大となるニックネームが複数ある場合には、全てのニックネームを生成結果として出力する。

2.4 統合手法

本手法では、規則に基づく手法の生成結果と辞書検索に基づく手法の生成結果を統合することでニックネームの生成を行う。両手法の生成結果からより良いニックネームを選択し生成結果とする。

ニックネームの選択は、日本名と生成したニックネームの類似度を利用する。辞書検索に基づく手法のときと同様に、類似度は、日本名と生成したニックネームとの先頭からのマッチ文字数と定義する。両手法から生成された各ニックネームに対して、類似度を計算し、最大となるものを最終的な生成結果として出力する。ただし、以前と同様の理由により、類似度を計算する際に、付録に示す正規化規則でニックネームの正規化を行う。また、規則に基づく手法については、日本名のベースに該当する部分の母音連続（**VVC**）が存在する場合は後の母音を削除する処理も行う。最終的に、類似度が最大になる全てのニックネームを生成結果として出力する。

3. 評価実験

対象データは、兵庫県内のX小学校3年生2クラス65名分の名前とした。名前の重複が6件存在したため、名前の異なり数は62であった。また、英語名と英語ニックネームの辞書として、Webデータ（<http://usefulenglish.ru/vocabulary/>）の“Men's Names and Nicknames”と“Women's Names and Nicknames”を用いた。

これらのデータを使用して、規則に基づく手法、辞書検索に基づく手法、統合手法のそれぞれでニックネームを生成した。ただし、辞書検索に基づく手法で、複数のニックネームが生成される場合は、最大で3件のニックネームをランダムに選択し生成結果とした。また、この結果と規則に基づく手法を統合したものを統合手法の生成結果とした。

各生成結果に対して性能評価を行った。主観評価により、生成されたニックネームの適切／不適切を判断した。判断は、第一著者と第二著者が独立に行った。二人の判断が一致しないニックネームについては、相談の上、最終的な判断を下した（不一致の件数：規則に基づく手法4件、辞書検索に基づく手法17件、統合手法5件）。

評価尺度は、Recall, Precision, *F*-measureの3種類を用いた。それぞれ、

$$R = \frac{\text{適切なニックネームが生成できた日本名の数}}{\text{日本名の数}}, \quad (1)$$

$$P = \frac{\text{適切なニックネームの数}}{\text{生成したニックネームの数}}, \quad (2)$$

$$F = \frac{2RP}{R+P}, \quad (3)$$

で定義した。

実験結果を表1に示す。表1より、規則に基づく手法が、全体性能（*F*-measure）では、一番性能が良いことがわかる。一方、辞書検索に基づく手法の性能は非常に低い。統合手法の性能は、Recall重視の生成であることがわかる。

表1: 実験結果

Method	<i>R</i>	<i>P</i>	<i>F</i>
規則に基づく手法	0.846	0.965	0.902
辞書検索に基づく手法	0.323	0.347	0.335
統合手法	0.877	0.791	0.832

4. 考察

上述の通り、全体性能としては、規則に基づく手法が最も良い結果となった。規則に基づく手法では、日本名、英語名、英語ニックネームのデータの分析から得られた規則を利用するため、精度よくニックネームの生成が行える。また、Recallも高いことから、大部分の日本名に対応できることもわかる。少数ではあるが、規則に基づく手法で失敗するケースとして、ベースが取得できない場合が存在する（62種類の日本名に対して、ベースが取得できないケースは7件存在した）。例えば、“Noa”のような母音連続で終わる日本名が挙げられる。英語名では、母音連続で終わることは希であり、英語名のデータを基本とする提案手法では対象外となっている。ベースの決定方法を拡張することで、このケースに対応することができると予想される。また、直接的な解決策ではないが、名付けデータベース^(注4)を利用する方法も考えられる。名付けデータベースとは、名前の意味、起源、発音などの情報が収録されたデータベースのことである。“Noa”のように比較的短い日本名かつベースが取得できない場合は、日本名と共に名付けデータベースから得られた情報を提示する。この場合、日本名とニックネームが同一になってしまうが、付加情報があるため児童が不満に感じることは少ないと予想される。例えば、

(注4): 例えば、<http://babynamesworld.parentsconnect.com/>や<http://www.thinkbabynames.com/>などが利用可能である。

児童: 私の名前は、乃愛 (のあ) です。

ロボット: Hi. Your name means “from love”. Can I call you Noa?

のような会話を考えることができる。この例では、下線部が名付けデータベースから取得した情報(名前の意味)である。

表 1 からわかるように、辞書検索に基づく手法の性能は非常に低い。この理由として、日本名と英語名で綴りのシステムが異なることが挙げられる。例えば、“Yuu (ゆう)” のような、英語では希である綴りでは、結果が得られないことが多い。この問題の対処法として、正規化規則を拡張することが考えられる。また、別の方法として、発音で類似度を定義することも考えられる。文字ではなく発音に基づいて類似度を定義すれば、音が似たニックネームが生成できる。

予想に反して、統合手法の全体性能は、規則に基づく手法より低くなった。この一番の理由として、辞書検索に基づく手法の性能が非常に低いことが挙げられる。辞書検索に基づく手法の性能が低いため、規則に基づく手法と統合しても性能が改善しないと分析できる。また、規則に基づく手法も辞書検索に基づく手法も、同じデータ(英語名と英語ニックネームのデータ)を基礎としているため、両手法とも同じ日本名に対して生成に失敗する傾向が見られた。そのため、辞書検索に基づく手法で、規則に基づく手法を補うことができたのは僅か 2 件であった。

全体性能では規則に基づく手法に劣るが、統合手法にも、バリエーションという点で利点がある。規則に基づく手法では、一つの名前に対して生成できるニックネームは一種類のみである。一方、統合手法では、規則に基づいた手法と辞書検索に基づいた手法を統合するため、複数のニックネームが生成できる可能性がある。実際、実験結果を分析したところ、正しく生成できたニックネームの一人あたりの数は、規則に基づいた手法では 0.88、統合手法では 1.0 であった。また、65 名の名前に対して、重複なくニックネームを割り当てることができた割合を求めたところ、規則に基づいた手法では 0.71、統合手法では 0.74 であった。したがって、同名の児童が存在する場面では、統合手法のほうに利点があるといえる。

5. おわりに

本稿では、児童と英語学習支援ロボットとの円滑なコミュニケーションを目的として、英語のニックネームを生成する手法を提案した。具体的には、規則に基づく手法、辞書検索に基づく手法、統合手法の 3 種類を提案した。実験の結果、全体性能としては、規則に基づく手法が最も良いことが明らかとなった。また、生成できるバリエーションという観点からは、統合手法に利点があることも明らかとなった。

今後は、正規化規則と生成規則を拡張することにより、生成性能の改善を行う予定である。また、辞書検索に基づく手

法において、発音により類似度を定義することで性能改善を行う予定である。更に、生成されたニックネームを利用した、会話や学習活動の生成に取り組む予定である。

参考文献

- [1] C.W. Chang, J.H. Lee, P.Y. Chao, C.Y. Wang and G.D. Chen, “Exploring the Possibility of Using Humanoid Robots as Instructional Tools for Teaching a Second Language in Primary School,” Educational Technology and Society, vol.13, no.2, pp.13–24, 2010.
- [2] C. Graham, Creating Chants and Songs, A. Maley, ed., Oxford University Press, Oxford, 2006
- [3] 猪井新一, “英語活動に関する小学校教員の意識調査,” 茨城大学教育実践記要, vol.28, pp.49–63, 2009.
- [4] 神田崇行, 佐藤瑠美, 才脇直樹, 石黒浩, “対話型ロボットによる小学校での長期相互作用の試み,” ヒューマンインタフェース学会論文誌, vol.7, no.1, pp.27–37, 2005.
- [5] S. Lee, H. Noh, J. Lee, K. Lee and G.G. Lee, “Cognitive Effects of Robot-Assisted Language Learning on Oral Skills,” Proc. Interspeech 2010 Workshop on Second Language Studies: Acquisition, Learning, Education and Technology, 2010.
- [6] 寄田明宏, 久保田直行, “第 2 言語学習のためのロボット援用教育,” 第 25 回人工知能学会全国大会, no.3B1-OS22c-3, pp.1–4, 2011.

付 録

日本名の正規化のための規則: 日本名のローマ字表記のバリエーションを正規化するための規則である。例えば、最初の規則は, “shi” は “si” に統一することを意味する。

shi → si
chi → ti
tsu → tu
fu → hu

英語名と英語ニックネームの正規化のための規則: 英語名および英語ニックネームにおけるアルファベットをローマ字のアルファベットに正規化するための規則である。

ca → ka
cu → ka
ci → si
ce → se
co → ko
ph → f
l → r
q → k
v → b

ニックネーム整形用の規則: 規則に基づく手法で生成したニックネームを整形するための規則である。

hh → ff
jj → zz