

日本語-ウイグル語辞書の自動作成とその収録語の分析

ムフタル・マフスット† 小川 泰弘‡ 杉野 花津江‡ 稲垣 康善‡

†名古屋大学大学院国際開発研究科 ‡名古屋大学大学院工学研究科

著者らは、既存のウイグル語-日本語辞書を基にして、見出し語数約2万の日本語-ウイグル語辞書を半自動的に作成した。この辞書が日常よく使われる語彙をどの程度含んでいるかなどの特性を調べるために、国立国語研究所の教育基本語彙6,104語のうちより基本的とされている日本語基本語彙2,071語、及びEDR日本語テキストコーパスの出現頻度上位2,056語に対し、日本語-ウイグル語辞書のカバー率を調査し、いずれについても約80%のカバー率であることが分かった。カバーされなかった単語について、逐一その理由を調べ、判明した種々の理由を整理すると共に、それに基づいて分類をした。

その結果、辞書作成をする時にカバー率を上げるために注意すべき点などについていくつかの知見を得ることができた。本稿では、それらについて述べる。

Automatic Generation of Japanese-Uighur Dictionary and Its Vocabulary Analysis

Muhtar Mahsut† Yasuhiro OGAWA‡ Kazue SUGINO‡ Yasuyoshi INAGAKI‡

†Graduate School of International Development, Nagoya University ‡Graduate School of Engineering, Nagoya University

{muhtar, yasuihiro, sugino, inagaki}@inagaki.nuie.nagoya-u.ac.jp

The authors have constructed automatically a Japanese-Uighur dictionary consisting of about 20,000 items. This paper describes the process of generating our Japanese-Uighur dictionary from an available Uighur-Japanese one. We have investigated the vocabulary of our dictionary and found it includes about 80% of 2,000 high-priority words of Japanese. We have also investigated the reasons why each word of the remaining 20% of 2,000 high-priority words is not included in our dictionary, and have classified the words not included to five groups according to the reasons we found through our investigation.

1 はじめに

著者らは、実用に近い日本語-ウイグル語機械翻訳システムの実現を目指して一連の研究をしてきており [1, 2, 3, 4, 5, 6]、その過程で、一定の語彙数を持つ日本語-ウイグル語電子辞書の開発が不可欠であると考え、その開発に着手した。その時点では、日本語-ウイグル語に関する通常の辞書さえない状況であった。最初は、基本的な日本語-ウイグル語機械翻訳実験用の辞書の開発を考えて、IPAの計算機用日本語基本動詞辞書 IPAL [8] をベースに、名詞や形容詞などを含め、約1200語の日本語-ウイグル語電子辞書を作成した [1, 2]。IPAL 動詞辞書には、日本語の動詞のうちで語彙体系上ならびに使用頻度上重要であると考えられる基本的な和語動詞861語が含まれている。両言語のなかで特に格助詞を含む名詞接尾辞と動詞接尾辞が動詞と密接な関係があり、日本語-ウイグル語機械翻訳においても、動詞が重要であるため、IPAL 動詞辞書を選んだ。しかし、1200

語前後の辞書では不十分であり、実用に近い機械翻訳システムの実現には、少なくとも日常使われる最低限の語彙を含む日本語-ウイグル語電子辞書の開発が必要であるとの考えに至った。

そこで、我々はまずウイグル語-日本語辞書であるウイグル語辞典 [7] を電子化して機械可読にし、その逆辞書を自動的に生成するという方針で日本語-ウイグル語電子辞書の開発に着手した [9, 14]。日本語-ウイグル語電子辞書の開発作業は次のような段階に分けて行なった。

1. ウイグル語-日本語電子辞書の作成。
 - 1-1. ウイグル語辞典 [7] のデータの電子化と項目タグの付与。
 - 1-2. 各項目の修正および品詞の付与。
2. 日本語-ウイグル語電子辞書の作成。
 - 2-1. ウイグル語-日本語辞書から日本語-ウイグル語辞書を自動生成。

2-2. 各見出し語の検査および修正 .

2-3. 機械翻訳システムで利用できるように変換 .

以上のような一連の作業の結果、語彙数約 20,000 語の日本語-ウイグル語電子辞書を作成することができた。我々は、この辞書が日常よく使われる語彙をどの程度含んでいるかを調べるために

a. 国立国語研究所の教育用基本語彙 [11]6,104 語中のより基本的とされている 2,071 語に対するカバー率

b. EDR コーパス [12] の日本語テキスト文に含まれる単語の上位頻度 2,056 語に対するカバー率

の 2 点に関して調査した。その結果、約 80% のカバー率であった。そして、a. 及び b. でカバーされなかった単語に関して、カバーされなかった理由を詳細に分析し、その理由を大きく A~E の 5 つに分類し、それぞれを更に細分類した。その結果、本論文と同様の手法で辞書を作成する時にカバー率を上げるために注意すべき点などについていくつかの知見を得ることができた。本論文では、それらについて述べる。

2 ウイグル語-日本語辞書の電子化

我々の目標は、日本語-ウイグル語機械翻訳システムに使用できる辞書の作成である。元となる日本語-ウイグル語辞書が存在しなかったため、まず文献 [10] の維漢辞典の和訳であるウイグル語辞典 [7] を計算機で処理可能な形にするための電算入力から作業を始めた。この作業では、計算機での処理を容易にするために、各項目にタグを付加し、図 1 のような形式でデータを入力した。ここで、\u、\j、\ue 及び\je は、それぞれウイグル語の語彙見出し、日本語の対訳語、ウイグル語の例文及びその和訳文を指示するタグである。日本語の対訳語が複数ある場合、意味的に近いものと、そうではないものが、それぞれ‘,’と‘;’で区切られ、例文が完成文である場合やその和訳文の末尾に‘.’が付いている。又、ウイグル語には 32 の文字があり、‘c’以外のローマ字 25 文字とウイグル語の発音に近いローマ字の前に‘!’を付した‘!g’, ‘!h’, ‘!k’, ‘!e’, ‘!o’, ‘!u’, ‘!z’ の 7 文字の合計 32 の表記に対応させた。

¹‘gilitserin’ はウイグル語でも「外来語」である。

次に、このデータに対し、ウイグル語ネイティブ話者および日本語ネイティブ話者からなる著者らの共同作業により、すべての見出し語とその項目について、その内容を吟味・確認し、更に以下の点に注目して修正・校正を施した。

- 1) 品詞の付与および、複数の品詞をもつ見出し語を品詞毎に分離
- 2) 不適切な訳語の修正
- 3) 語源・語義が異なる見出し語を異なる語義毎に分離

まず、1) であるが、図 1 に見るように、元のウイグル語辞典 [7] には品詞が付与されていなかったため、すべての見出し語に人手で品詞を付与した。また、複数の品詞を持っている見出し語については、品詞ごとに分けて別々の見出しとした。日本語と違って、ウイグル語では同じ単語が二つ以上の品詞を持つ場合が多く、特に、形容詞にも名詞にもなる語が多い。そのような場合に、別々の見出し語に分割し、日本語訳および例文も、それぞれの品詞ごとに分けて付加した。また、今回使用したウイグル語辞典 [7] は、維漢辞典 [10] の中国語訳を日本語に翻訳して作成されており、二次翻訳による意味のずれや欠落がかなりあった。2) は、主にそうした点に関しての修正である。例えば、ウイグル語の‘gilitserin’¹は、維漢辞典 [10] での中国語への訳語が“甘油”になっており、ウイグル語辞典 [7] でも日本語訳が「甘油」になっていた。しかし、「甘油」は、日本語国語辞書の見出し語として含まれていなかった。そこで、中国語-英語辞典 [13] を引くと、“甘油”の英訳が‘glycerine’であることが分かり、ウイグル語の‘gilitserin’に対して「グリセリン」という日本語訳語を当てた。3) は複数の異なる語義を持つ単語が一つの見出しになっている場合で、これを別々の見出しとした。元になった維漢辞典 [10] では一つの見出し語として扱われているものがウイグル語辞典 [7] では別々の見出し語になっていたり、またはその逆であったりして、同じ見出しにするか、別見出しにするかの基準がはっきりしていないので、我々のこの作業では、意味が類似していなければ、別の見出しとした。

こうした作業の結果、最終的に約 16,000 語の語彙量を持つウイグル語-日本語電子辞書を作成することができた。この辞書の品詞毎の見出し語数と平均対訳数を表 1 に示す。

\u n!etij!e \j 結果, 結論, 成果, 効果, 成績 \
 \ue musabi!kining n!etijisi \je 試合の結果.\
 \ue !uginix n!etijisi \je 学業成績.\
 \ue ~ !kazanma!k \je 成績を得る.\
 \ue n!etijig!e erixm!ek \je 成果を得る.\
 \ue ~ b!erm!ek \je 効果があがる.\
 \ue ~ qi!kma!k \je 結果が出る.\
 \ue ~ qi!karma!k \je 結論を出す.\

図 1: 電子化されたウイグル語辞典の一部

表 1:ウイグル語-日本語電子辞書の見出し数及び平均対訳数

品詞	名詞	動詞	形容詞	動作 名詞	副詞	助数詞	感嘆詞	代名詞	接続詞	接続 助詞	語気 助詞	合計
見出し数	7,259	3,682	2,868	1,046	691	142	48	19	21	10	2	15,788
日本語への 平均対訳 語数	1.86	2.35	2.12	2.12	2.09	1.61	2.14	1.79	2.48	2.80	2.50	2.05

3 日本語-ウイグル語電子辞書の半自動生成

前章で述べた一連の手続きによって作成したウイグル語-日本語電子辞書に対して、機械処理を施すことによって逆辞書である日本語-ウイグル語辞書を生成した。

ウイグル語-日本語電子辞書では、一つのウイグル語見出し語に複数の日本語訳が付加されている場合があるが、それらに対して、まず、これをウイグル語と日本語訳語の対にし、それにウイグル語-日本語電子辞書に記されていた品詞、および、漢字かな変換プログラム KAKASI²を利用しての読み仮名を加え、4 項組 < 日本語見出し語, 読み, 品詞, ウイグル語訳語 > を自動的に作成した。それらを更に読みでソートして、結果的に図 2 のような形で日本語-ウイグル語辞書の基本を作成した。さらに、この 4 項組に対し、1) 見出し語としての適切さ 2) ウイグル語訳の妥当性 3) 読みの正しさの 3 点に注意しながら、人手による修正・校正を加えた。これらの作業の結果、語彙数約 2 万の日本語-ウイグル語電子辞書を作ることができた。表 2 に、日本語-ウイグル語電子辞書の品詞毎の見出し語数及び平均対訳

語数を示す。この表の形容詞には、いわゆる日本語文法で言う形容詞、形容動詞、連体詞が含まれている。名詞、動詞、形容詞、サ変名詞の合計で 19,133 語となり、総見出し語数 20,166 語の 94.9% を占めている。また、日本語 1 単語あたりのウイグル語訳語数は、各品詞毎の平均で 1.31 ~ 1.76 語であり、全体の平均で 1.56 である。これは、本辞書の作成手続きからも推察されるように、ウイグル語-日本語辞書のウイグル語見出ししか、日本語-ウイグル語辞書の訳語に出現しないことから当然の結果とうなづかれよう。また、表 1 では、ウイグル語見出し語数 15,788 に対して日本語訳語数の平均が 2.05 であるから、 $15,788 \times 2.05 = 32,365.4 \approx 32,300$ のウイグル語-日本語訳の対があることになるのに対して、表 2 では、我々の作成した日本語-ウイグル語辞書には $20,166 \times 1.56 = 31,458.96 \approx 31,500$ の日本語-ウイグル語訳の対が存在していることを示している。この差の約 1000 の対の減少は、日本語-ウイグル語辞書の見出し語に換算すれば、 $1,000 \div 1.56 = 641.03 \approx 640$ 語に相当する。これらの減少は、上述の操作の 1), 2) に関連する作業で不適切であるとして、削除されたものである。

表 2:日本語-ウイグル語電子辞書の見出し数及び平均対訳数

品詞	名詞	動詞	形容詞	サ変名詞	副詞	助数詞	感嘆詞	接続詞	合計
見出し数	8,457	5,411	3,480	1,785	784	156	73	20	20166
ウイグル語への 平均対訳語数	1.51	1.60	1.68	1.31	1.76	1.47	1.71	1.55	1.56

²<http://kakasi.namazu.org/>

(日本語見出し語)	(読み)	(品詞)	(ウイグル語訳語)
はけ	はけ	名詞	qotka
はけでみがく	はけでみがく	他動詞	qotkilima!k
豚	ぶた	名詞	qox!ka
ブラシ	ぶらし	名詞	qotka
ブラシをかける	ぶらしをかける	他動詞	qotkilima!k

図 2: ウイグル語-日本語辞書から日本語-ウイグル語辞書への変換

4 日本語-ウイグル語電子辞書の収録語の分析

これまでに説明してきた一連の作業によって出来上がった日本語-ウイグル語電子辞書が、1 で述べた目標に対してどの程度満足できるか、どのような単語をその見出し語として含んでいるのか、どのような単語を含んでいなかったのかについて 2 つの調査を行なった。一つは、国立国語研究所が発表している教育基本語彙 [11] に対するカバー率であり、もう一つは EDR コーパス [12] の日本語テキスト文に含まれる出現頻度の高い単語に対するカバー率である。前者は日本語の基本語彙と認められている語に対するカバー率であり、後者は新聞記事や雑誌記事に高い頻度で出現する単語に対するカバー率である。まず、第 1 の調査について述べる。国立国語研究所 (以下では国語研と略記する) の「教育基本語彙データベース」[11] に含まれている 6,104(2,071+4,033) 語の内、どれぐらいが日本語-ウイグル語電子辞書の見出しに現れるかを調べた結果、3,552 個 (58.20%) の単語が含まれていることが分かった。ここで 2,071+4,033 と書いているのは、最も基礎的な基本語彙の 2,071 語、及びそれに続く基本語の 4,033 語と、国語研の語彙データベース中で 2 つに分けているので、そのように表している。

第 2 の調査に関しては、我々は約 21 万文の入っている EDR コーパスの日本語テキスト文中の形態素の中から出現頻度の上位 6,055 語³を選択し、その中でどれぐらいの数の単語が見出しに現れるかを調べた。その結果、3,135 個の単語 (51.77%) が我々が作成した日本語-ウイグル語電子辞書の見出しに出現した。

上で述べた調査方法 1,2 の結果は、ある意味では、よく使われている単語の約半分が見出し語の中に含

まれていないことを示している。これは基本語彙を 6000 語ベースで考えた時の結果である。我々は、国語研の最も基本語彙とされる 2000 語ベースで考え、それぞれ国語研の 2,070 語、EDR の上位頻度 2056 語に対して日本語-ウイグル語電子辞書のカバー率がどれぐらいになるかを調査した。その結果、国語研に対して 80.14%(1659 語)、EDR に対して 78.89%(1622 語) のカバー率であった。

ここで、我々は、カバーされなかった単語 (国語研 19.86%、EDR 21.11%) に関して、そのカバーされなかった理由や現象を調査し、国語研、及び EDR のそれぞれの単語に対して A, B, C, D, E の 5 種類の分類基準を決定し、その同一の基準に基づいて分類を行なった。その結果を表 3 に示す。次に、各グループについて詳細な考察を行なう。

A) 日本語-ウイグル語辞書に含まれると見なすことのできる単語; これらの単語は、表記の違いや品詞の違いによって辞書に含まれないと判断されたものがほとんどであり、それらに対して適当な機械処理を行なうことで辞書に含めることができる。従って、2000 語ベースのカバー率は実質的に国語研は 82.27%、EDR は 79.37%となり、国語研に対するカバー率は少し上である。A(1)-A(3) では、国語研と EDR との間で差が出ている。特に A(1) に分類される語が国語研で多かった理由は、第一に、国語研の表記に関するポリシーによるものであり、例えば、国語研の表記には、「然し・併し」や「然も・而も」のように通常使用しない漢字が多用されていたのが原因である。第二に、EDR の出典は新聞や雑誌が主であるため、漢字からなる単語が比較的が多いことなどによると考えられる。

B) 元のウイグル語-日本語辞書で日本語の訳付けが不十分と思われる単語; 元のウイグル語-日本語辞書 [7] において、日本語訳を付けた時に、ウイグル

³ここで、6,055 になっているのは基本語彙とされる国語研の「教育基本語彙データベース」の語彙数は 6,103 であり、それに準じて 6,000 前後の上位頻度語を選んだ結果である。

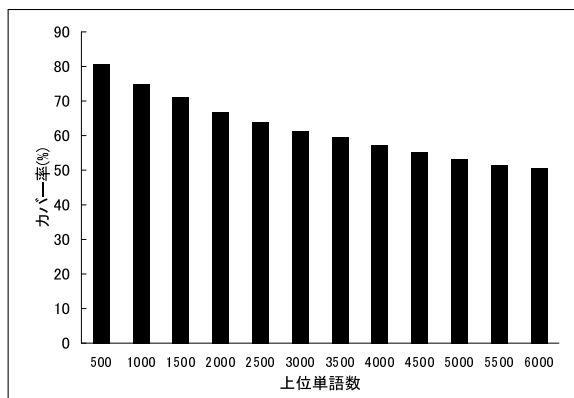


図 3: 上位頻度単語の日-ウ辞書に含まれる割合

語-中国語辞書 [10] を基にしたことにより、二次的な意味の欠落、あるいはウイグル語単語の意味が十分理解されなかったことで、これらの単語がウイグル語単語の訳語として辞書に現れなかったと思われる。また、元の辞書 [7] は、日本語話者を対象にしているためか、ウイグル語単語の日本語訳が複数考えられても、その中からいくつかだけを訳語にしている傾向がある。例えば、ウイグル語の単語 ‘hletlerlik’ の訳として「危険な」と「危ない」が考えられるが、日本語話者には、その一つがあれば、‘hletlerlik’ の意味を理解するのに十分であるため、「危ない」がその訳の候補から落ちている。従って、ウイグル語-日本語辞書の編集時に、できるだけ多くの訳語を入れるべきであることが分かる。種類 B に含まれる単語は、カバーされなかったものの内、国語研では約 2 分の 1、EDR では 3 分の 1 強を占めていた。これは国語研の基本語彙が和語が比較的多いのにに対し、EDR コーパスの文が漢字だけからなる単語を多用する新聞や雑誌から多く採られているという特徴を示していると言える。

C) ウイグル語訳が単純な一語で表せないもの; これらの単語は、ウイグル語訳が派生語であるか、二つ以上の語の組み合わせからなるため、元のウイグル語辞書に表れていなかった。種類 C に属する単語のような現象は、日本語-ウイグル語のみならず、他の言語間の辞書作成時にも当然現れる。これらの単語は、人手によって辞書に登録する必要がある。

D) そのウイグル語訳が元のウイグル語-日本語辞書 [7] を作った時点でまだ単語として使われていないか、何らかの理由で見出しから外れた単語; このグループの単語は、国語研と EDR ではかなりの差がある。国語研では技術用語が 1 個しかないのに対

し、EDR では 26 個であり、二つの文献の性格の違いがよく現れている。種類 D の単語は辞書に新たに登録する必要がある。

E) 日本語単語が表す概念がウイグル語で複合語や句で表すことのできないもの; グループ E のように、一方の言語の 1 単語が表す概念を他方の言語の単語で簡潔に表現できない問題は、どの言語間でも一定の割合で存在する。この類の単語が国語研に関して比較的多く見られているのは、出典が新聞や雑誌である EDR と比べて日本語特有の概念が国語研の基本語彙に多いためであると考えられる。これらの単語は現段階ではどうしようもないが、辞書に例文を多く列挙したり、説明を付けたりすることによって対処することになる。

上で行なった考察から、我々がウイグル語-日本語辞書から半自動的に作成した日本語-ウイグル語辞書は、基本用語 2000 語ベースで考えた時、国語研、EDR とともに 80% 前後のカバー率が実現されている。また、出現頻度が高いほど、カバー率が高いことも我々の調査で明らかになっている (図 3)。これらの結果は、人間が普通の辞書として使用する観点から見れば、辞書の質が一定以上に保持されていると言える。

5 まとめ

本論文では、我々が作成した日本語-ウイグル語電子辞書の作成方法と、そのカバー率、及びカバーされなかった単語についての調査結果を報告した。その結果から見て、実用にかなり近い辞書が作成できたと言える。

今回の辞書作成とその分析から、次のようなことが言える。

- 1) 逆辞書を生成する前に、オリジナルの辞書の単語の訳語を拡充する必要がある。
- 2) オリジナルの辞書の例文に出てくる複合語を辞書の見出し語として追加すべきである。
- 3) 逆辞書を生成する前に、オリジナルの辞書の見出し語に品詞を付与するだけでなく、その訳語にも品詞を付与すべきである。

今後語彙のさらなる整備を行ない、実用レベルの日本語-ウイグル語機械翻訳に利用したいと考えている。

表 3:カバーされなかった単語の調査結果

種類	現象	例	対処方法	
A	辞書に含まれると見なすことのできる単語 国語研: 44 語 EDR: 10 語	(1) 表記の違いや送り仮名の違いによるもの。 国語研: 35 語 EDR: 9 語	「何時も」、「贈り物」は日本語-ウイグル語辞書の見出しに入っていないが、「いつも」、「贈物」はその見出しに入っている。	これらの単語に対して一定の機械的処理を行うことにより、辞書に追加できると考えられる。
		(2) 日本語で、形容詞の語幹が形容動詞の語幹にもなるもの。 国語研: 3 語 EDR: 0 語	ウイグル語の形容詞 'illi!k' の訳語として「暖かい」は辞書にあるが「暖かな」はない。 	
		(3) 日本語の動詞で、その意味がウイグル語では形容詞で表されるもの、または逆に、日本語の形容詞で、その意味がウイグル語では動詞で表されるもの。 国語研: 6 語 EDR: 1 語	「優れる」は、動詞であるがそのウイグル語訳である 'artu!k' は形容詞である。 「痛い」は、形容詞であるがそのウイグル語訳である 'algrima!k' は動詞である。	
B	元のウイグル語-日本語辞書で、日本語の訳付が不十分と思われる単語 国語研: 205 語 EDR: 168 語	(1) 同じ概念の違った表現 国語研: 26 語 EDR: 15 語	「米国」と「アメリカ」, 「火曜」と「火曜日」など。	ウイグル語-日本語辞書の編集時にできるだけ多くの訳語を入れる。
		(2) その単語の類似語が辞書に存在する 国語研: 58 語 EDR: 42 語	「辺り」に対して「周り」, 「危ない」に対して「危険な」など。	
		(3) 元のウイグル語-日本語辞書でウイグル語に語義が複数あるのに、その一部が落ちている。 国語研: 78 語 EDR: 95 語		
		(4) ウイグル語-日本語辞書の編集時に、複数の品詞を持つウイグル語に対して品詞の付け忘れがあったもの。 国語研: 28 語 EDR: 15 語	ウイグル語の形容詞の「赤い」(!kizil) は名詞にも成り得るが、[名詞] という品詞を付けなかった。 	
		(5) 日本語の丁寧語、尊敬語、謙譲語などの待遇表現 国語研: 15 語 EDR: 1 語	「参る」、「おっしゃる」, 「いただく」など。	

表 3:カバーされなかった単語の調査結果 (続き)

	種類	現象	例	対処方法
C	ウイグル語訳が単純な一語で表せないもの。 国語研: 151 語 EDR: 191 語	(1) ウイグル語訳が派生語であるもの。 国語研: 9 語 EDR: 28 語	従来 → !esli-dle	辞書に新たに登録する。
		(2) ウイグル語訳が 2 語以上の組合せからなる複合語であるもの。 国語研: 130 語 EDR: 150 語	彼処 → awu y!er あの ところ	
		(3) その語をウイグル語に訳する時、言い換えや語順の変更が必要になるもの。 国語研: 11 語 EDR: 10 語	... 過ぎ → !h!eddidin oxulk ... (限度を超えて...)	
		(4) その日本語を含む連語をウイグル語に訳すもの。 国語研: 1 語 EDR: 3 語	しれ → かもしれない → ... m!umkin	
D	そのウイグル語訳が元のウイグル語-日本語辞書を作った時点でまだ単語として使われていないか、何らかの理由で見出しから外れた単語 国語研: 8 語 EDR: 66 語	(1) 技術的用語。元の辞書ができた時点で、まだ使われていなかった用語。 国語研: 1 語 EDR: 26 語	「テレビ」、「データ」など。	辞書に新たに登録する。
		(2) ごく普通に使われるウイグル語であるが、元のウイグル語-日本語辞書の見出し語から外れたもの。 国語研: 4 語 EDR: 1 語	「人形 (!koqa!k)」、「もしもし (w!ey)」など。	
		(3) 固有名詞。 国語研: 0 語 EDR: 31 語	「リクルート」、「大阪」など。	
		(4) ウイグル語では接尾辞として訳されるもの。 国語研: 3 語 EDR: 8 語	「... 的」、「... 性」、「... 目」など。	

表 3:カバーされなかった単語の調査結果 (続き)

	種類	現象	例	対処方法
E	その概念がウイグル語で表すことのできないもの。 国語研: 23 語 EDR: 9 語	(1) その概念が元々ウイグル語にない 国語研: 18 語 EDR: 6 語	「昭和」、「量」、「神社」など。	
		(2) 総称としての概念が、ウイグル語にはないもの。 国語研: 3 語 EDR: 1 語	「親」、「親子」、「菓子」など。	
		(3) 対応する語法がウイグル語には存在しないもの。 国語研: 2 語 EDR: 2 語	「お」(接頭語)、「ご」(接頭語)。	
合計	国語研: 411 語 EDR: 434 語			

参考文献

- [1] ムフタル・マフスット, 外山勝彦, 稲垣康善: 日本語とウイグル語の助詞の対応関係とその機械翻訳への応用, 電子情報通信学会言語理解とコミュニケーション研究会技術報告, NLC93-60(1993).
- [2] ムフタル・マフスット: 日本語-ウイグル語機械翻訳に関する研究, Master's thesis, 名古屋大学大学院工学研究科情報工学専攻 (1993).
- [3] ムフタル・マフスット, 外山勝彦, 稲垣康善: 日本語-ウイグル語機械翻訳における助動詞のパラメータ化による処理, 電子情報通信学会言語理解とコミュニケーション研究会技術報告, NLC94-13(1994).
- [4] M. Mahsut, F. Casablanca, K. Toyama and Y. Inagaki: Particle-Based Machine Translation for Altaic Languages: the Japanese-Uighur Case, *Proc. of the 3rd Pacific Rim International Conference on Artificial Intelligence*, Vol.2, pp.725-731(1994).
- [5] ムフタル・マフスット, 小川泰弘, 稲垣康善: 日本語-ウイグル語機械翻訳のための格助詞の変換処理, 自然言語処理, Vol.8, No.3, pp.123-142 (2001).
- [6] 小川泰弘, ムフタル・マフスット, 杉野花津江, 外山勝彦, 稲垣康善: 日本語-ウイグル語機械翻訳における派生文法に基づくウイグル語動詞句の生成, 自然言語処理, Vol.7, No.3, pp.57-77 (2000).
- [7] 飯沼英三: ウイグル語辞典, 穂高書店 (1992).
- [8] 計算機用日本語基本動詞辞書 IPAL(Basic Verbs), 情報処理振興事業協会技術センター (1987).
- [9] ムフタル・マフスット, 杉野花津江, 小川泰弘, 稲垣康善: 日本語-ウイグル語電子辞書の開発, 電気関係学会東海支部連合大会講演論文集, No.624(1996).
- [10] 新疆大学中文系編: 維漢辞典, 新疆人民出版社 (1982).
- [11] 国立国語研究所: 教育基本語彙の基本的研究-教育基本語彙データベースの作成-, 国立国語研究所報告 117, 明治書院 (2001).
- [12] 日本電子化辞書研究所: EDR 電子化辞書 1.5 版, EDR 日本語コーパス (1996).
- [13] 朱原ら編: 精選英漢・漢英辞典, 商務印書館・Oxford University Press(1986).
- [14] 小川泰弘, 杉野花津江, ムフタル・マフスット, 稲垣康善: 日本語-ウイグル語電子化辞書の自動生成, 電気関係学会東海支部連合大会講演論文集, No.568(1998).