

翻訳メモリと機械翻訳の融合に向けて ～ 用例翻訳 S D M T によるアプローチ ～

加藤 直人

N H K 放送技術研究所

katou.n-ga@nhk.or.jp

1 はじめに

翻訳支援には翻訳メモリと機械翻訳による方法が考えられる。翻訳メモリは過去の翻訳用例を検索するという現実的な翻訳支援技術である。一方、機械翻訳は人手を介さずに自動的に翻訳するという理想的な技術ではあるが、現在はその精度がまだ十分ではない。さらに充実した翻訳支援を考えると、機械翻訳と翻訳メモリという2つ方法を密接に連携させることがことが好ましいこと考えられる[1][2]。そこで、本稿では用例翻訳 S D M T [3][4]を利用することにより、翻訳メモリと機械翻訳を融合することを提案する。一例として災害・気象ニュースの日英翻訳について述べる。

2 翻訳メモリ：翻訳用例提示システム

翻訳メモリとは原言語と目的言語間の用例をデータベースとして保存しておき、翻訳者によるキーワードに基づいてそのデータベースに格納される対訳には、単語、句（節）、文、文章のレベルがある。

翻訳メモリによる翻訳支援では、翻訳作業の主体は翻訳者であり、翻訳メモリはそれを助ける手段となる。翻訳者は過去の翻訳例を参照することができ、効率的に高品質な翻訳をすることが可能となる。

N H K でも翻訳メモリによる翻訳支援として、翻訳用例提示システムを開発している[5]。このシステムはデータベースとして、過去に翻訳された N H K ニュースの蓄積しており（例えば、日本語や英語の場合には現在までのところ過去10年分ほどの量である）翻訳者は過去の翻訳例やそれまでの報道経緯を検索することができる。

図1に日本語と英語のデータベースを使った場合の翻訳用例提示システムを示す。図の左側が日本語のニュース原稿、右側が人手で翻訳された英語原稿である。対訳は文章全体が表示されることにより、ある表現が使われた文脈を参照したり、日本語にはなく英語で追加された記述を知ることできる。現在では翻訳用例提示システムの多言語化も進めている[6]。

199704130033

台風一号発生

1997年04月13日10時52分

印刷

マーシャル諸島の西の太平洋上でごく小さく弱い台風一号が発生しました。気象庁の観測によりますと台風一号はきょう午前九時現在、マーシャル諸島の西のおよそ七百キロの太平洋上において中心の気圧が九百九十六ヘクトパスカル、中心付近の最大風速は十八メートルで中心から七十キロ以内では風速十五メートル以上の強い風が吹いています。台風は一時間におよそ十キロの速さで北北西へ進んでいて日本列島に直接の影響はありませんが、気象庁では近くを航行する船舶は台風の動きに注意するよう呼び掛けています。

199704130016 (ja: 199704130033)

16=TYPHOON-NO 1

社会、台風1号発生

1997/04/13/12:30

印刷

A small and weak typhoon has appeared in waters west of the Marshall Islands in the Pacific.

Japan's Meteorological Agency said on Sunday that the typhoon, the season's first, was located about 700 kilometers west of the Marshall Islands at nine a.m. Japan time.

The agency added that the typhoon has an atmospheric pressure of 996 hectopascals at its center.

It has a wind velocity of up to 65 kilometers per hour near its center and more than 54 kilometers per hour within a radius of 70 kilometers.

The agency said the typhoon is heading north-northwest at ten kilometers per hour and called on ships in the area to pay attention to the path of the typhoon.

図1 翻訳用例提示システム（日本語・英語）の検索結果

3 機械翻訳：S D M T

機械翻訳とは、事前に構築された翻訳知識に基づき、自動的に入力文を自動的に翻訳するものである。大量の文章を短時間のうちに翻訳してくれるとう大変便利なものであるが、現在は翻訳の質がまだ十分ではない。したがって、自動的に翻訳してくれるとはいうものの、実際には前編集や後編集など翻訳者による作業が必要となる。

機械翻訳の方式は、規則ベース翻訳(規則翻訳)、統計的機械翻訳(統計翻訳)、用例ベース翻訳(用例翻訳)の3つに大別される。規則翻訳では、翻訳知識は人手で構築され、非常に細かい言語情報を付与する必要がある。さらにその翻訳過程は非常に複雑である。したがって、ユーザはもとより開発者でさえ、なぜそのような翻訳結果が出てきたのかが不明な場合もある。統計翻訳では、翻訳知識は自動獲得され言語情報も規則翻訳ほどは細くないものの、やはり翻訳過程はブラックボックスとなっている。一方、用例翻訳では用例そのものが翻訳知識であり、翻訳する際にはこの用例を直接利用するので翻訳過程の透明性は比較的高い。

本稿では機械翻訳として、用例翻訳の1つのアプローチであるS D M Tを採用した。S D M Tでは、後述するように、翻訳で利用した用例を介することで、翻訳メモリと機械翻訳を融合することができる。以下では、このS D M Tについて基本的な考え

方を説明する。詳細については文献[4]を参照されたい。

S D M Tでは入力に類似した用例を組み合わせるにより翻訳を行う。用例翻訳で中心となるのは類似度であるが、S D M Tでは単言語内類似度と二言語間類似度に基づいて翻訳を行う。単言語内類似度は用例収集のための類似度であり、用例翻訳では一般的に使われているものである。一方、二言語間類似度は、2つの言語間にまたがる類似度であり、本手法を特徴付けるものである。

図2の例を使って説明しよう。入力文 J_0 が与えられると、S D M Tは単言語内類似度により、入力文に類似した用例を収集する。今の例では、3つの用例が収集されており、強調して表示された単語が入力文に一致した単語である。構文的に類似した文を収集するためには機能語重視の検索を行い、内容的に類似した文を収集するためには内容語重視の検索を行う。出力文は用例の英語側の単語をうまく選択し組み合わせることで生成されるが、それには言語モデルと二言語間類似度を制約としている。

S D M Tは原言語側での処理と目的言語側での処理が分かれており、両言語にわたる言語変換処理がないという特徴がある。また、二言語間のおける句や単語のアライメント情報を使えば、さらなる精度の向上が期待できる。

【入力文】

J_0 : 気象庁の観測によりますと、フィリピンのミンダナオ島の東の太平洋で熱帯低気圧が台風1号になりました。

【用例1】単言語内類似度 = 0.89

J_1 : 気象庁の観測によりますと、太平洋のフィリピンの東の沖合で熱帯低気圧が台風1号になりました。

E_1 : The Meteorological Agency says the first typhoon of this year has formed in the east of the Philippines.

【用例2】単言語内類似度 = 0.89

J_2 : 気象庁の観測によりますと、フィリピンのルソン島の東の海上で熱帯低気圧が台風26号になりました。

E_2 : The Meteorological Agency says the 26th typhoon of this year has formed in the east of Luzon Island in the Philippines.

【用例3】単言語内類似度 = 0.71

J_3 : 気象庁によりますとフィリピンのミンダナオ島の東の太平洋で台風が発生しました。

E_3 : Weather officials say a typhoon has appeared in the Pacific far east of Mindanao island in the Philippines.

【出力文】

E_0 : The Meteorological Agency says the first typhoon of this year has formed in the Pacific far east of Mindanao Island in the Philippine.

図2 入力文に類似した用例

3 翻訳メモリと機械翻訳の融合

最近の機械翻訳ソフトの多くが翻訳メモリ機能を備えてはいるが、機械翻訳手法は規則翻訳によるものであり、機械翻訳と翻訳メモリとの融合が密とは言い難い。

以下では、現在研究を進めている気象・災害ニュースの日英翻訳を例にとり、翻訳メモリと機械翻訳の融合について説明する。気象・災害ニュースは、使われる語彙がある程度限定でき、また定型的な表現も多く、比較的機械翻訳しやすい分野でもある。

提案するシステムでは、日本語のニュース原稿が与えられると、まずSDMTにより文単位で翻訳を行い、結果を表示する。例えば、出力のイメージは図3のようになる。左側の日本語が翻訳したい文(入力文)であり、右側がSDMTによる翻訳結果である。この例ではすべての日本語文が翻訳されているが、類似した用例がない場合(実際には、単言語内類似度がある閾値以下の用例しかない場合)には、翻訳結果は出力しない(空欄のままとなる)。

翻訳者は各翻訳結果を見て、翻訳結果が妥当のものであればそれを使い、そうでなければ、その修正を行ったり、ゼロから入力文を翻訳する。

翻訳結果を修正する際に参考にできるのが、入力文に類似した過去の翻訳用例である。提案システムでは、そのような翻訳用例としてSDMTで利用したものを提示する。例えば、図4のように、**S** ボタンをクリックするとSDMTで利用した3つの用例が表示される。このとき入力文や出力文に一致している単語が強調されて表示される。このような翻訳用例は、翻訳結果が妥当である場合でもこれらを参照することにより、翻訳者はその翻訳結果の信頼性を高めることができるようになる。

さらに、その用例上の**◀** ボタンをクリックすると、その用例文を含むニュース全体を表示することができる。例えば、図4では、SDMTで利用した1番目の用例を含むニュースを表示している。ニュース全体を表示することで、文脈情報や追加記述も参照することが可能となる。すなわち、従来の翻訳用例提示システムと同様の機能を備えてい

る。また、機械翻訳は文単位で翻訳を行うため、その文の翻訳としては問題がないものの、他の文と並べたときに文章全体としての収まりが悪い場合がある。このような場合にもニュース全体の用例を参照することにより、それを修正することが可能である。

同様に、**P** ボタンや**W** ボタンをクリックすることにより、それぞれ、SDMTの翻訳に実際に利用された、句や単語の情報を見ることができる。日本語側についているボタンは英語側と同じ情報を参照できるが、表示される言語が左右逆転している。

以上のように、SDMTで利用した、文、句、単語の用例を介することにより、翻訳メモリと機械翻訳を融合している。翻訳メモリで提示される用例は機械翻訳でも利用されているので、非常に密な融合となっている。

4 おわりに

用例翻訳SDMTによる、翻訳メモリと機械翻訳の融合について述べた。今後は機械翻訳の精度向上を行いながら、翻訳メモリを含むをトータルなシステム作りを行う。そして、気象・災害ニュースを対象とした実用的なシステムを構築し、将来的には他の言語や分野にも拡張していきたい。

参考文献

- [1]潮田明ほか．機械翻訳と訳例検索を統合した翻訳支援システム．言語処理学会第9回年次大会，pp.413-416，2003．
- [2]潮田明ほか．自動翻訳から翻訳支援へ，そして…．情報処理学会誌，Vol.44，No.9，pp.931-939，2003．
- [3]加藤直人．マルチプルアライメントによる用例翻訳．第4回科学技術フォーラムFIT2005，pp.175-178，2005．
- [4]加藤直人．SDMT：用例翻訳への新しいアプローチ．情報処理学会自然言語処理研究会，NL-170，pp.151-156，2005．
- [5]熊野正ほか．翻訳用例提示システムの設計・開発・運用，電子情報通信学会論文誌，Vol.J84-D-11，No.6，pp.1175-1184，2001．
- [6]Goto, Isao et al．A multi-language translation example browser. Procs. of MT Summit IX, pp.463-4666, 2003．

<p>200605093504 台風1号が発生 2006年05月09日22時14分</p>	
<p>S 今夜、フィリピンの東の太平洋で台風一号が発生しました。 P W</p>	<p>The first typhoon of this year has formed in the Pacific far east of Philippine. S P W</p>
<p>S 気象庁の観測によりますと、フィリピンのミンダナオ島の東の太平洋で熱帯低気圧が台風一号になりました。 P W</p>	<p>The Meteorological Agency says No. 1 typhoon of this year has formed in the Pacific far east of Mindanao Island in Philippine. S P W</p>
<p>S 中心の気圧は千ヘクトパスカル中心付近の最大風速は十八メートルで中心から半径二百二十キロ以内では風速十五メートル以上の強い風が吹いています。 P W</p>	<p>The agency added that the typhoon has an atmospheric pressure of 1000 hectopascals at its center. S P W It has a wind velocity of up to 65 kilometers per hour near its center and more than 54 kilometers per hour within a radius of 220 kilometers.</p>
<p>S 台風は一時間におよそ十キロの速さで西へ進んでおり、気象庁は近くを通る船舶は注意するよう呼びかけています。 P W</p>	<p>The agency said the typhoon is heading west at ten kilometers per hour and called on ships in the area to pay attention to the path of the typhoon. S P W</p>

図3 機械翻訳による翻訳結果

<p>200605093504 台風1号が発生 2006年05月09日22時14分</p>	
<p>S 今夜、フィリピンの東の太平洋で...</p>	<p>The Meteorological Agency says the first typhoon of this year has formed in the east of Philippines.</p>
<p>200110130169 台風1号発生 2001/10/13 04:29</p>	<p>The Meteorological Agency says the first typhoon of this year has formed in the east of Philippines.</p>
<p>The first typhoon has formed in the Pacific far east of Philippine.</p>	<p>気象庁の観測によりますと、太平洋のフィリピンの東の沖合で熱帯低気圧が台風1号になりました。</p>
<p>The Meteorological Agency says the first typhoon of this year has formed in ...</p>	<p>気象庁の観測によりますと、フィリピンのルソン島の東の海上で熱帯低気圧が台風26号になりました。</p>
<p>The typhoon has an atmospheric pressure of 996 hectopascals at its center ...</p>	<p>気象庁によりますとフィリピンのミンダナオ島の東の太平洋で台風が発生しました。</p>
<p>The Agency says the typhoon is almost stationary ...</p>	<p>The agency said the typhoon is heading west at ten kilometers per hour and called on ships in the area to pay attention to the path of the typhoon.</p>

図4 翻訳メモリによる検索結果