

5Z-07 英文作成支援のための対訳付き用例検索ツール TWIX

後藤啓次 早川哲史 兵藤安昭 池田尚志
岐阜大学工学部

1 はじめに

英文を書くにあたり、英語の用例を見て作成することがある。しかし、紙に印刷された通常の冊子体の用例辞書では、見出し語となっている単語でしか引くことができず、また、たくさんの用例の中から目的に合ったものを探すのが大変である。本研究では、全用例を対象にした全文検索による辞書引きができること、ANDとORのブール演算による絞り込みに加えて、係り受け共起関係を利用した絞り込みができること等を実現した用例検索ツールを開発した。なお、対象として用いた辞書は、オーム社発行、富井篤編「科学技術英大辞典」である。

2 システムの概要

本システムの構成を図1に示す。辞書データは、9つに分類されており、これを図2に示すようなデータベースに取り込む。本システムには検索方法が二つある。全用例を対象にした全文検索と通常の辞書引きのように見出し語で検索する見出し語検索である。全文検索に関しては、まず初めに、辞書中の日本語用例文に対して我々の研究室で開発している日本語解析システムIBUKIを用いて形態素・構文解析を行った[1]。次に、単語の出現情報や係り先情報からなるインデックスデータを作成し、係り受け関係等を用いた全文検索を行っている。また、用例中の係り受け関係のみを抽出したテーブルを作成しており、用例検索インターフェイス上で用例の絞り込みの際に利用することができる。係り受け関係を用いた全文検索システムの内容については[2]を参照されたい。見出し語検索に関しては、辞書テーブル上の見出し語、読みなどのフィールドを直接検索している。

Japanese-English Parallel Corpus Retrieval System
TWIX as an English Writing Support Tool
K. Goto, S. Hayakawa, Y. Hyodo, T. Ikeda
Faculty of Engineering, Gifu University
Gifu-shi,501-1193,Japan

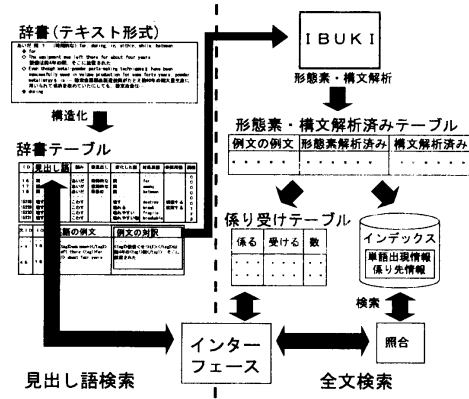


図 1: システムの構成図

3 検索インターフェイス

全文検索を絞り込む条件として以下の4つを指定することができる。

- OR 条件** 入力した語を少なくとも1つ含む用例を検索する
- AND 条件** 入力した語すべてを含む用例を検索する
- 近接条件** 入力した語の距離がある範囲以内の用例を検索する
- 係り受け条件** 入力した語が係り受け関係で出現する用例を検索する

辞書(テキスト形式)

あいだ 間 1. (時間的な) for, during, in, within, while, between
◆ for
◇ The equipment was left there for about four years
設備は約4年の間、そこに放置された。
◇ Even though metal-powder parts-making techniques § have been successfully used in volume production for some forty years, powder metallurgy § is ... 粉末金属部品製造技術がたとえ約40年の間大量生産に用いられて成功を取っていたにしても、粉末冶金は...
◆ during

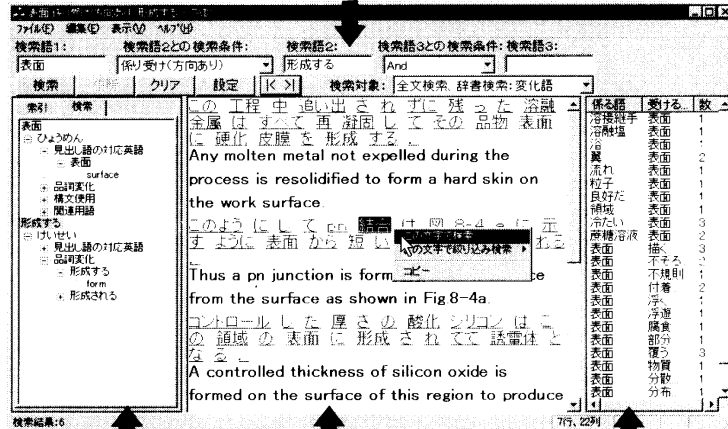
辞書テーブル

ID	見出し語	読み	節見出し	変化した語	対応英語	参照用語	属性
1 6	間	あいだ	時間的な	間	for		0
2 3	間	あいだ	空間的な	間	among		0
2 4	間	あいだ	事象の	間	between		0
15228	壊す	こわす		壊す	destroy	破壊する	0
15229	壊す	こわす		壊れる	break	故障する	1
15230	壊す	こわす		壊れやすい	fragile		2
15231	壊す	こわす		壊れやすい物	breakable		3

文ID	ID	英語の例文	例文の対訳
4 4	1 6	The <tag2>equipment</tag2> was left there <tag1>for </tag1> about four years.	<tag2>設備<せつび></tag2>は約4年の<tag1>間</tag1>、そこに放置された。
4 5	1 6		

図 2: 辞書 RDB

検索語の入力



辞書見出し 対訳付き用例 係り受け関係表示

図 3: 用例検索インターフェース

インターフェイスは、Visual C++を用いて作成した。インターフェイスのウィンドウは、検索語入力ウィンドウ、辞書見出し表示ウィンドウ、対訳付き用例表示ウィンドウ、係り受け関係表示ウィンドウの4つからなる(図3)。検索語を入力して検索ボタンを押すと検索した用例を用例表示ウィンドウに表示し、検索語に関して係る語、及び受ける語を係り受け関係表示ウィンドウに表示する。また、検索された日本語用例は単語ごとにリンクが張っており、クリックするとその単語に関連して絞り込み検索を行うことができる。また同様に、係り受け関係表示ウィンドウで係る語、受ける語を見て、クリックすることでも絞り込み検索を行うことができる。

3.1 検索例

例えば「表面を形成する」という用例を検索したい場合、「表面」と「形成する」をAND条件で検索すると次のような用例が得られたとする。

1. 外側の表面は平らで内側の薄板は…が通過できる浅いくぼみを形成するように作られている。
2. エピタクシャル層の表面を酸化させ元素を形成させる区域以外の部分から酸化物を選択的に除去する。
3. この工程中追いつけられずに残った熔融金属はすべて再凝固してその品物表面に硬化皮膜を形成する。

AND検索では、1, 2のように「表面」と「形成する」が係り受け関係にないものも検索されてしまうが、本システムでは係り受け関係を指定することで3のみを絞り込むことができる。

4 おわりに

英文作成支援のための対訳付き用例検索ツールを開発した。このシステムでは従来の見出し語検索以外にも、全用例を対象にした全文検索による辞書引きができ、ANDとORのブール演算による絞り込みに加えて、係り受け共起関係を利用した絞り込みができる。またインターフェイス上の単語をクリックすることで、検索結果に対する絞り込みをすることができる。今後は、使用経験を踏まえて、より使いやすいインターフェイス、検索方法を検討していく予定である。

最後に、電子化データを利用して頂いたオーム社に感謝いたします。

参考文献

- [1] 兵藤安昭, 若田光敏, 池田尚志: あいまいさを許すロバストな係り受け解析システム(言語処理学第5回年次大会ワークショップ論文集, 言語処理学会, pp23-28, 1999)
- [2] 新美和彦, 兵藤安昭, 池田尚志: 係り受け関係を用いる高精度全文検索(情報処理学会第55回全国大会講演論文集, 4N-10, 1997)