

B3ゼミ

仮名漢字変換

1. 仮名漢字変換

仮名漢字変換...文節単位で行われる。

例.「はしらない」

1.「はしらない」が文節であるとして形態素解析した場合

はしら 動詞、ラ行五段活用、未然形、語幹「走」

ない助動詞「ない」、終止形

↓

「はし」の部分だけが漢字に変換可能

結果「走らない」

1. 仮名漢字変換

2. 辞書だけを用いる場合

単語辞書に掲載されている単語

1字 は:葉、齒、派、刃 2字 はし:橋

し:市、詩

ら: 3字 はしら:柱

な:名、菜

い:意

「はしら」が「はしる」の未然形であることが分からない。

1. 仮名漢字変換

ヒューリスティクス・・・経験的な法則

ヒューリスティクス1

日本語では一字語が並ぶことは少ない。

「葉氏ら菜意」などが削除される。

ヒューリスティクス2

漢字に変換される部分は文節の文頭の自立語のみ。

「箸ら菜い」、「端らな井」などが削除

1. 仮名漢字変換

ヒューリスティクスを適用すると

端らない、橋らない、箸らない、柱ない

辞書を用いた場合

変換候補多数

活用する語の変換では正解が出ない場合がある

同音語

「ほしょう」→「保証、保障、補償、歩哨」

コンピュータがこの中から自動選択することは非常に困難。

「こうしょう」→「交渉、校章」

「こうしょうする」のように「する」が付いたときは区別可能

共起辞書の利用

共起・・・関連の深い語は文の中で共に生起しやすい

「なつはあつい」

「夏」「暑い」 共起する

「夏」「熱い」 共起しない

ただし共起関係はヒューリスティクスであり、正しいことが保証されているわけではない。

文脈に依存する同音語の選択

「あついてっぱん」→「熱い鉄板」、「厚い鉄板」

どちらも意味的には正しい

- 文脈情報を用いれば判別可能
- 語のニューラルネットワーク

大統領は法案が通過すれば直ちに公布するつもりである。しこうに

「しこう」→思考、施行、指向

法案、公布→施行である可能性が高い