

機械翻訳のための 対訳フレーズアプローチ

池田秀人
(有)サイバープロ

翻訳のためにはどんな辞書が必要か

- ◎ 単語辞書(品詞付き、出現頻度付き)
- ◎ 熟語辞書(品詞付き、出現頻度付き)
- ◎ シソーラス(同義語、反対語、派生語、関連語、上位語、分類)
- ◎ 概念辞書(階層構造、言語独立性)
- ◎ 構文規則辞書(品詞列規則集)
- ◎ 共起辞書
- ◎ 文型辞書
- ◎ 言換えデータベース
- ◎ 対訳コーパス(翻訳メモリー)
- ◎ 対訳フレーズ辞書

新しい辞書

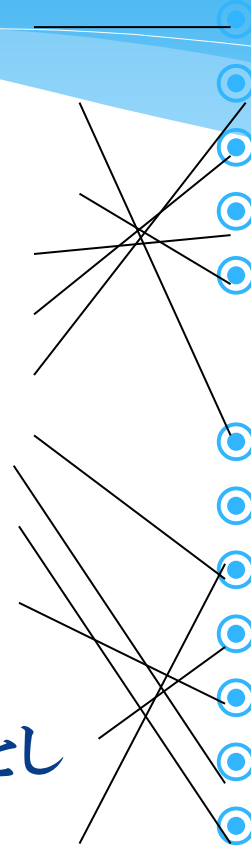
線形フレイズ(LP)

- ◎ 近年、地球温暖化などの地球環境保全問題からクリーンエネルギーへの期待が高まっており、太陽光を電気エネルギーに直接変換する太陽電池は、クリーンなエネルギー源として注目されている。
- ◎ 近年、/地球温暖化などの/地球環境保全問題から/クリーンエネルギーへの/期待が/高まっており、/太陽光を/電気エネルギーに//直接変換する/太陽電池は、/クリーンなエネルギー源として/注目されている。/

LP2による完全対訳例

- ◎ 近年、
- ◎ 地球温暖化などの
- ◎ 地球環境保全問題から
- ◎ クリーンエネルギーへの
- ◎ 期待が
- ◎ 高まっており、
- ◎ 太陽光を
- ◎ 電気エネルギーに
- ◎ 直接変換する
- ◎ 太陽電池は、
- ◎ クリーンなエネルギー源として
- ◎ 注目されている。

- ◎ In recent years,
- ◎ there have been increasing
- ◎ expectations
- ◎ for clean energy
- ◎ in light of the global
- ◎ environment protection
- ◎ problems
- ◎ such as global warming.
- ◎ A solar battery
- ◎ has received attention/
- ◎ as a clean energy source
- ◎ because it directly converts
- ◎ sunlight
- ◎ into electric energy.



線形フレーズ辞書は有効か

- * 完全フレーズ対応はいつでも可能か
- * フレーズ翻訳は一意か
- * フレーズの文脈自由翻訳が可能か
- * すべての線形フレーズを集めたら、いくつあるか
- * フレーズの定義は？()
- * フレーズの連結規則は？(文法)

NO!

NO!

NO!

数百万

文中の連続した意味のある語列？

埋めこみ規則！

構文階層フレーズ(SHP)

- ◎ S0=近年、_([S1])
- ◎ N2=_など([N3])
- ◎ N3=地球温暖化
- ◎ N4=_の_([N2],[N5])
- ◎ N5=地球環境保全問題
- ◎ C6=_から([N4])
- ◎ N7=クリーンエネルギー
- ◎ C8=_への([N7])
- ◎ N9=期待
- ◎ C10=_が(N9])
- ◎ V11=高まっており
- ◎ N12=太陽光
- ◎ C13=_を([N12])
- ◎ N14=電気エネルギー
- ◎ C15=_に([N14])
- ◎ V16=変換する
- ◎ V17=直接_([V17])
- ◎ N18=太陽電池
- ◎ C19=_は、([N26])
- ◎ N20=エネルギー源
- ◎ N21=クリーンな_([N20])
- ◎ C22=_として([N21])
- ◎ V23=注目されている。
- ◎ S24=___([C6],[C8],[V11])
- ◎ S25=___([C15],[V17])
- ◎ N26=___([S25],[N18])
- ◎ S27=___([C22],[V23])
- ◎ S1=_、_([S24],[S27])

依存階層フレーズ(DHP)

- ◎ S0=近年、_
◎ N2=地球温暖化
- ◎ N3=_の_
◎ N4=地球環境保全問題
- ◎ N5=クリーンエネルギー
- ◎ N6=_への期待
- ◎ P7= から が高まる
([N3],[N6])
- ◎ N8=太陽光
- ◎ N9=電気エネルギー
- ◎ P10=_を_に*変換する
([N8],[N10])
- ◎ P11=直接*([P10])
- ◎ N12=太陽電池
- ◎ N13=エネルギー源
- ◎ N14=クリーンな_
([N13])
- ◎ P15=@連用タ形:ており、_と
して注目されている。
([P7],[P15])

依存階層対訳フレーズ(DHP2)

- ◎ So=近年、_。 ([S1])
 - ◎ S1= から への期待が高まっております、_は、_として注目されている。
([N2],[N3],[P8],[N9],[N5])
 - ◎ N2= _などの_ ([N6],[N7])
 - ◎ N3=クリーンエネルギー
 - ◎ N6=地球温暖化
 - ◎ N7=地球環境保全問題
 - ◎ P8= _を直接_に*変換-する-
([N10],[N11])
 - ◎ N10=太陽光
 - ◎ N11=電気エネルギー
 - ◎ N9=太陽電池
 - ◎ N5=クリーンなエネルギー源
- ◎ So=In recent years, _([S1]).
 - ◎ S1=there have been increasing expectations for _ in light of . _ has received attention as _ because it _([N3],[N2],[N9],[N5],[P8])
 - ◎ N2= _ such as _([N6],[N7])
 - ◎ N3=clean energy
 - ◎ N6=global warming
 - ◎ N7=the global environment protection problems
 - ◎ P8=directly convert _ into _([N10],[N11])
 - ◎ N10=sunlight
 - ◎ N11=electric energy
 - ◎ N9=a solar battery
 - ◎ N5=a clean energy source

依存階層一意対訳フレーズ(DHUP2)

- * P= を直接 に*変換-する-([N1],[N2])
 - * P=directly convert into ([N1],[N2])
 - * P=changes into directly ([N1],[N2])
- * P= 太陽光を直接電気エネルギーに*変換-する-()
 - * P=directly convert sunlight into electric energy()
- * 引数依存対訳がある場合は、引数を関数部に移動

フレーズの種類と数

- * フレーズの種類
 - * 名詞フレーズ
 - * 動詞句フレーズ
 - * 文フレーズ
- * フレーズ数(分野非限定)
 - * 名詞フレーズ : 数百万
 - * 動詞句フレーズ: 数百万
 - * 文フレーズ: 数百万

80%は
自動的に抽出できる

現代のコンピュータなら
処理に問題はない

DHUP2を使って文入力支援

- * 翻訳品質の保証された文の入力
- * 知らない言語の文編集
- * 推奨文の作成
- * 翻訳者のノウハウ(文としてのまとめ)
- * 後編集支援

連絡先

- * (有)サイバープロ
- * 代表取締役 池田秀人
- * 〒525-0011 滋賀県草津市片岡町292-5
- * 077(568)1981 / 080(3850)7142

- * 注)まだWeb公開していません。