Research Project

フレーム意味論・構文的アプローチによるオンライン日本語語彙情報資源の構築

研究代表者：小原京子（理工学部）

日本語フレームネット 2002年7月～

日本語基本動詞の意味記述 → オンライン語彙情報資源化

生の日本語用例＜コーパス＞に基づく

背景知識＜フレーム＞に考慮

コーパスとは?

実際に話されたり、書かれたりした言葉のデータベース。
コーパスを使って辞書作りをすると辞書編纂者が自分の直観のみに頼るのではなく、客観的かつ豊富なデータに基づき、語の意味、文法パターン、他の語との関連などを記述できる。

コーパスを使った辞書作り

例）学習者用英英辞典COBUILD

The Bank Of Englishに基づく
40億語
話し言葉と書き言葉の両方
1990年以降のデータ
イギリス英語のみならずアメリカ英語等も

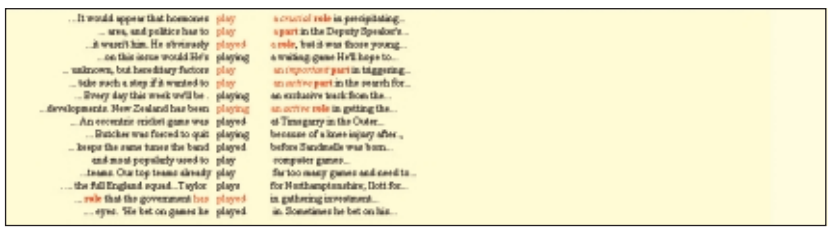
コーパスデータの検索

KWIC（Key Word In Context）コンコーダンス・プログラムを用いて検索

KWICコンコーダンスとは:
ある語がどのような文脈で用いられているかを示すもの。
特に、検索対象であるキーワードを中心に、その左右の文脈を合わせて表示したもの。

コーパスデータからわかること

“play” をキーワードとした場合の検索結果



“play” が “role” または “part” としばしば共起すること、さらに、“role” や “part” の前に “important” かそれに近い意味の形容詞が来ることがコンコーダンスからわかる。

日本語の辞書作りは遅れている?!

コーパスを使っていない市販国語辞典の記述例

【またがる】（跨る）
①股を開いて乗る。「鞍に―・る」
②一方から他方へかかる。わたる。「両方の分野に―・る研究」

この記述だと、「またがる」は二つの領域に関わる場合にしか用いられない、と解釈されるが…

↓

実際は、以下のコーパスデータが示すように「またがる」は三つ以上の領域についても用いられる。

(1)多摩ニュータウンは、八王子、多摩、町田、稲城の四市にまたがる三千ヘクタールに約十六万人が暮らす国内最大の住宅都市。

(2)最先端の研究を進めるには独創性のほかに、さまざまな分野にまたがる幅広い知識が必要になっている。
（「CD-毎日新聞95年版」より）

↓

従って、日本語の辞書作りにもコーパスは有益!

日本語フレームネットのコーパス

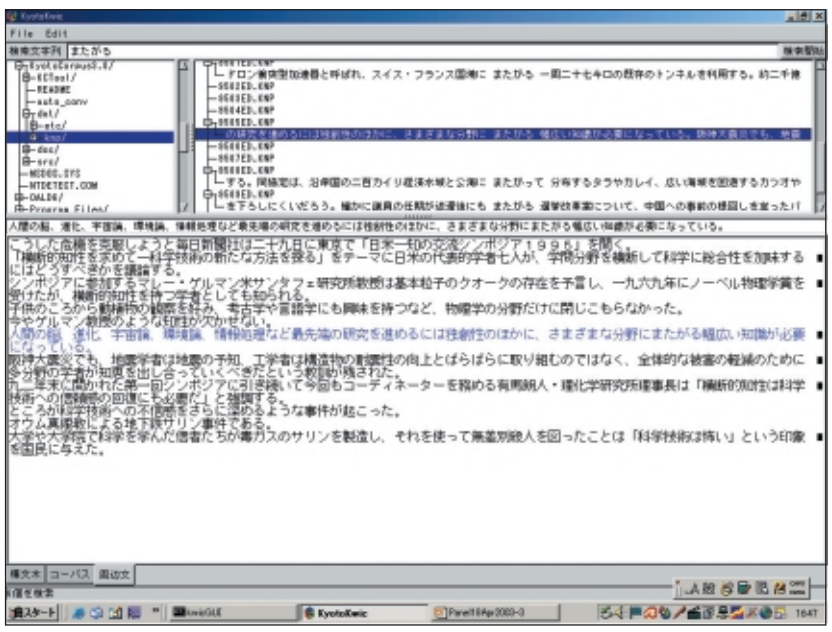
現　在

京都大学テキストコーパス、「CD毎日新聞95年版」から約40,000文収録

さらに

新聞記事、英日パラレルテキストを中心に拡充予定

日本語フレームネットのコンコーダンス・プログラム



共同研究者:石崎俊（SFC）大堀壽夫（東京大学）斎藤博昭（理工学部）鈴木亮子（経済学部）藤井聖子（東京大学）