

共起辞書

係り受け辞書

2021年12月22日

国分芳宏

次の文を例にして構文解析上での係先について考えてみましょう。

「ネットで行く公園を調べる。」

私はこんな悪文は書かないという人もいますが、このような文をときどき見受けます。「公園へ行く」が先に頭に浮かんでしまったのでしょうか。

この文を文節に分解すると

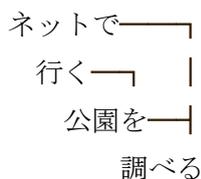
	始まりの用語	後ろへの影響
ネットで	体言で始まる	連用修飾書格
行く	用言で始まる	連体修飾書格
公園を	体言で始まる	連用修飾書格
調べる	用言で始まる	文末

ここで「ネットで」という文節は連用修飾格です。係り先として「行く」と「調べる」という2つの用言で始まる文が考えられます。

ネットで 行く

ネットで 調べる

よくある構文解析では距離的に近いので「ネットで」が直後の「行く」という文節に係ることになってしまいます。しかしこの係り受けは奇妙に感じます。ふさわしい組み合わせかどうかを調べるために、よくある組み合わせを集めた共起辞書を作りました。共起辞書には「ネットで」と「調べる」という組み合わせが登録してありますが、「ネットで」と「行く」は登録してありません。「ネットで」という文節は「調べる」という文節に係ることにします。図にすると。



ちなみに「ネットで」という文節を「車で」という文節に変えてみると、共起辞書には「車で 調べる」という組み合わせはないので「車で」直後の用言で始まる文節にかかることに

なり、構文構造は次のようになります。

車で―
行く―
公園を―
調べる

共起辞書

一文中に共起する用語の組み合わせを集めた辞書です。係り受け辞書とも言います。

用語の組み合わせで共起した組み合わせを全て集めると数が多すぎて爆発してしまいます。強い関係で文節を飛び越しても係りそうな組み合わせを人の手で集めました。

単独の用語では持たなかったのにネガポジの性質を共起辞書になると持つ性質があります。次の例では「消える」という用語にはネガティブ・ポジティブの性質は持ちませんが、共起辞書になると持つ組み合わせがあります。

資金が	消える	ネガティブ
あざが	消える	ポジティブ

多義語の区別ができる組み合わせがあります。「甘い」という用語には「甘味」という意味と「手ぬるい」という意味があります。この違いも係る用語を調べれば区別できます。

リンゴが	甘い	甘味
検査が	甘い	手ぬるい

また組み合わせた結果、意志動詞・無意志動詞の性質を持つようになるものもあります。

海を	望む	無意志
幸せを	望む	意思

筆者らの共起辞書には品詞、活用形、ネガティブ・ポジティブのフラグと意志動詞・無意志動詞のフラグが振ってあります。このフラグにより単独の用語では決められなかった多義語の違いを区別することができました。

2つの係り受けではすまされない3つ組み、4つ組みの共起語があります。この場合は意味的に切りやすいところで2つに分けました。

転んでも	ただでは起きない
爪に	火をともし
義を見てせざるは	勇なきなり
涙が	乾く暇もない

共起辞書に係受けが逆にした項目も採択しました。

暖かい 部屋

部屋が 暖かい

これまでに作った慣用句辞書などから約 91,000 語採択しました。

表記の揺れ

なるべくたくさんの方のケースで使えるように、よく使われると思われる用語を集めました。次の 2 冊の用語集を参考にして、用語の標準化を考えました。

記者ハンドブック 新聞用字用語辞典 共同通信社 第 1 3 版

読売新聞用字用語の手引き 第 6 版

しかし 2 つの辞典の間でも違いがあり最終的には筆者の判断で決めざるをえませんでした。特にひらがなで書くべきか、漢字を当てるべきかで悩みました。

将来の構想

辞書が大きくなります。参考文献ではシソーラスを組み込んであるので次に述べる仕事はすんでいますが将来次のような機能を付ける必要があります。

同義語の展開

「果物を食べる」という文では「果物」の同意語の「フルーツ」でも共起関係を定義できるようにする。

「フルーツを食べる」

さらに「果物」の狭義語である「リンゴ」「ミカン」などでも共起関係を定義できます。

「リンゴを食べる」

「ミカンを食べる」

またさらに辞書に意味のフラグを振っておく必要があるのですが、係り受けを意味で定義できるようにします。

地名へ 行く

という係り受けのために「東京」「京都」「奈良」・・・と、全ての地名との係り受けを作ることは不可能です。

参考文献

国分芳宏 シソーラスを組み込んだ意味解析システム 自然言語処理 Vol.17 No4