

係り受けの数理

著者	浅原 正幸
発行年	2021-09-10
URL	http://doi.org/10.15084/00003612



日本語の係り受け構造を数理的に考える

日本語の係り受け関係を有向グラフ上の全域木の推定問題に帰着する

日本語の係り受けには良い性質があり、係る・係らないの推定を線形回数行うことで推定できる

日本語の係り受け

文節単位に考える

日本語は文節単位に係り受けを考えると以下のような性質を持つ
主語（が・は・も）は述語（動詞・形容詞・形容動詞）に係る
目的語（に・を）は述語（動詞・形容詞・形容動詞）に係る
修飾語（説明する語）は被修飾語（説明される語）に係る
活用形は、係り先の品詞を決めることがある

グラフ構造としての係り受け

文節を頂点、係り受け関係を辺とした有向グラフとみなせる
係り受け構造の推定はグラフの全域木を推定する問題に帰着する
各辺にコスト（係りやすさ）を付与した場合、その総和を最小にする最小全域木問題として解くことができる

一方、日本語の文節係り受け構造は以下のような性質がある：

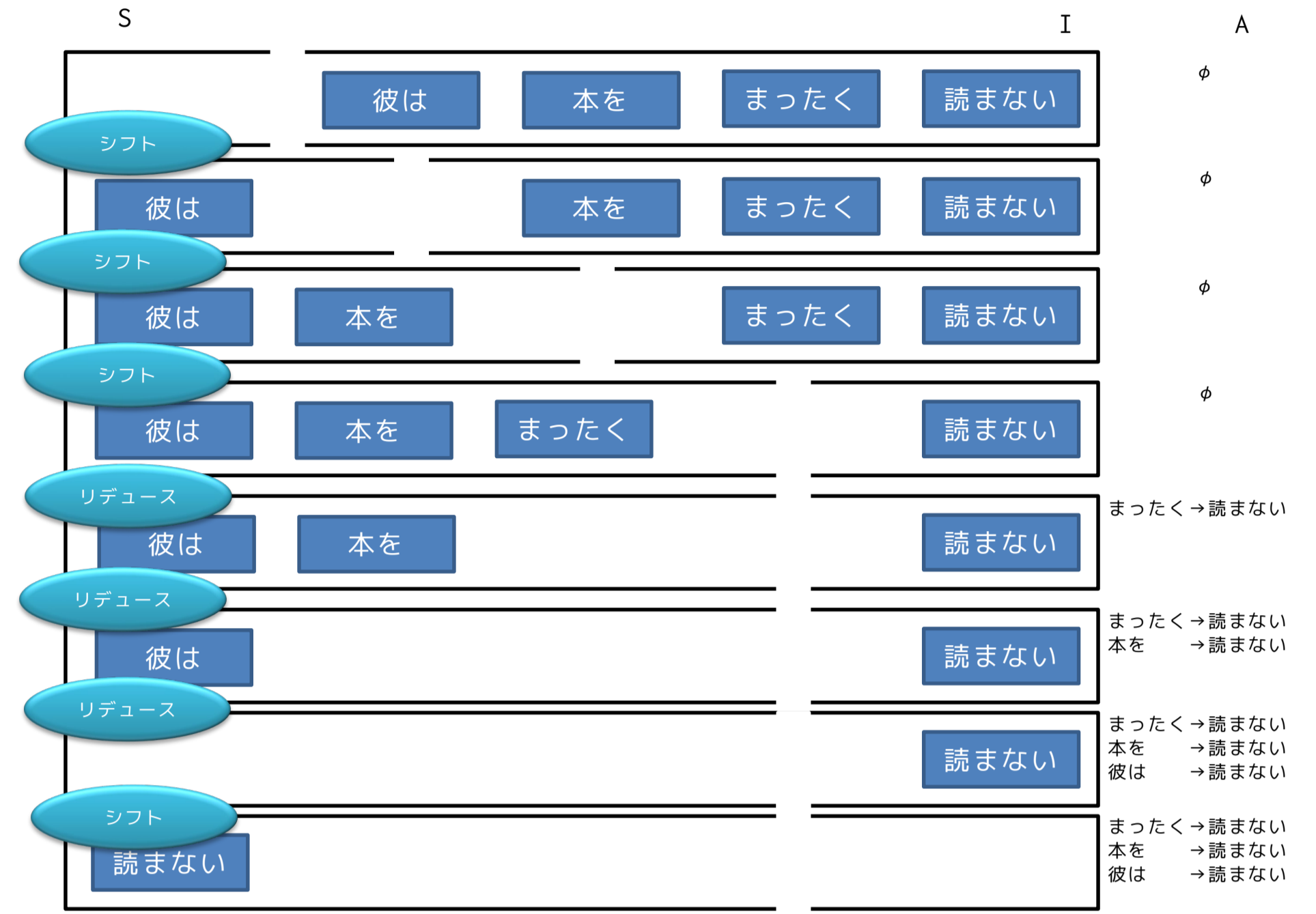
- ・常に左から右に係る (strictly head final)
- ・倒置や並列・同格構造を除いて基本的に係り受けは左から右
- ・平面グラフ上で非交差 (projective)
- ・係り受け木を平面に表現した際に、辺は交差しない

この性質により、shift-reduce 法により、文節の数 N に対して、高々 $2N$ の操作で文節係り受けを同定することができる

[Nivre 2003, Yamada and Matsumoto 2003]

Shift reduce 法による日本語係り受け解析

スタックとキューを準備することにより状態遷移し解析



入力文 $W = \langle w_1, w_2, w_3, \dots \rangle$
 状態: $\langle S, I, A \rangle$
 S: スタック I: キュー A: 係り受けの集合

状態遷移:
 初期状態 $\langle nil, W, \phi \rangle$
 終了状態 $\langle root, nil, a \rangle$

シフト: (mがnに係らない)
 $\langle m|S, n|I, a \rangle \rightarrow \langle n|m|S, I, a \rangle$

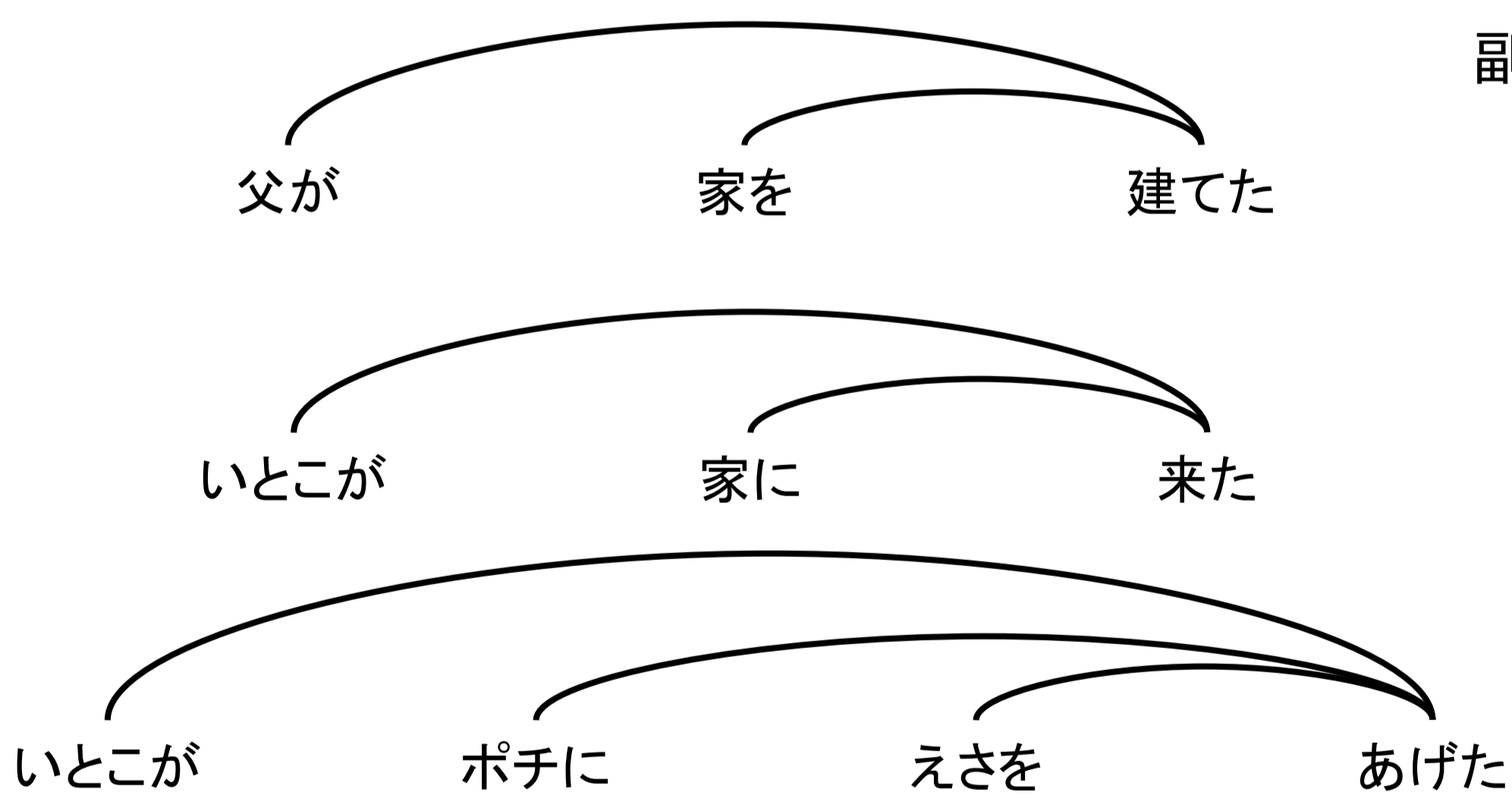
リデュース: (mがnに係る)
 $\langle m|S, n|I, a \rangle \rightarrow \langle S, n|I, a + 'm \rightarrow n' \rangle$

詳しくは
動画で

いろいろな係り受け関係

主語と述語と目的語

主語は述語に係る
目的語は述語に係る

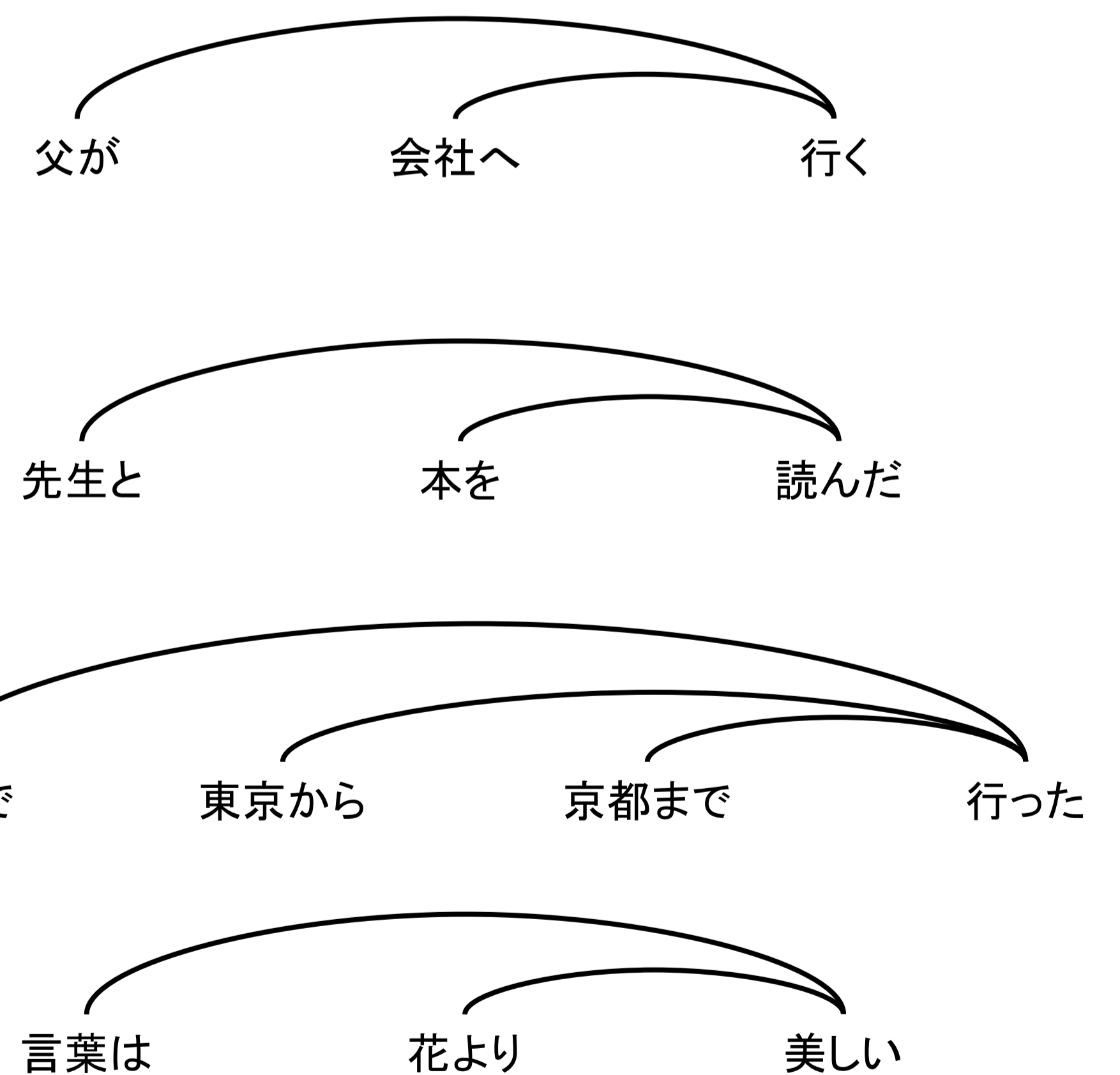


格助詞

が/を/に/へ/で/と/から/まで/より
格助詞を含む文節も述語に係る

副助詞

は/も
主題を表す副助詞も述語に係る



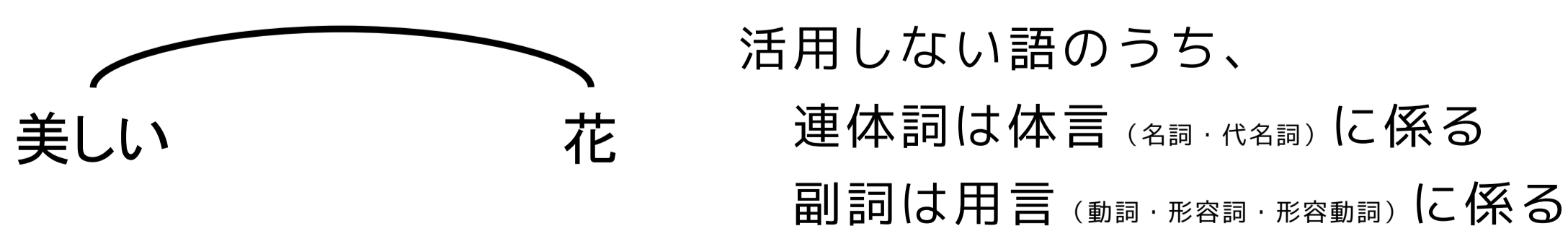
修飾-被修飾の関係

連体修飾

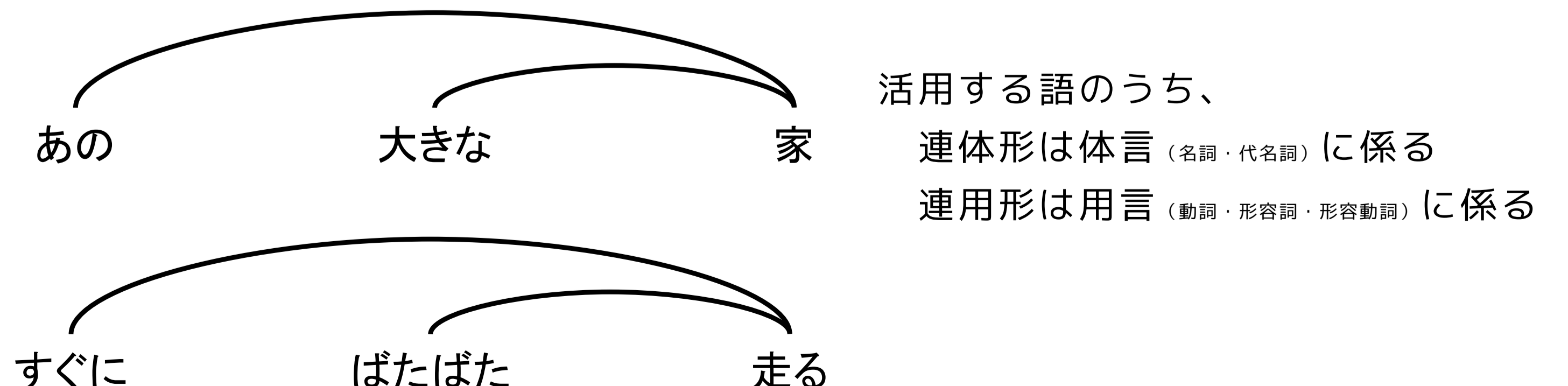
体言（名詞・代名詞）を修飾

連用修飾

用言（動詞・形容詞・形容動詞）を修飾



活用しない語のうち、
連体詞は体言（名詞・代名詞）に係る
副詞は用言（動詞・形容詞・形容動詞）に係る



活用する語のうち、
連体形は体言（名詞・代名詞）に係る
連用形は用言（動詞・形容詞・形容動詞）に係る