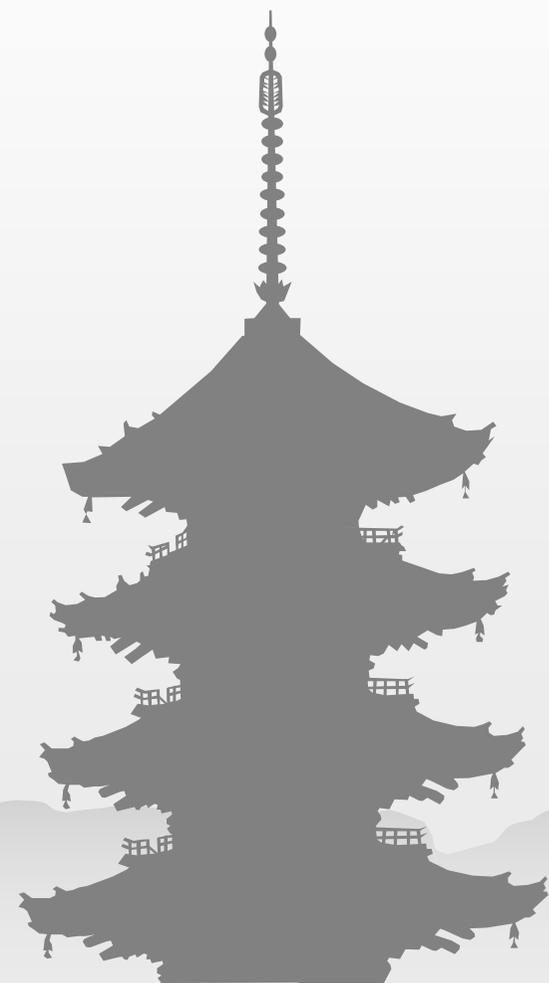
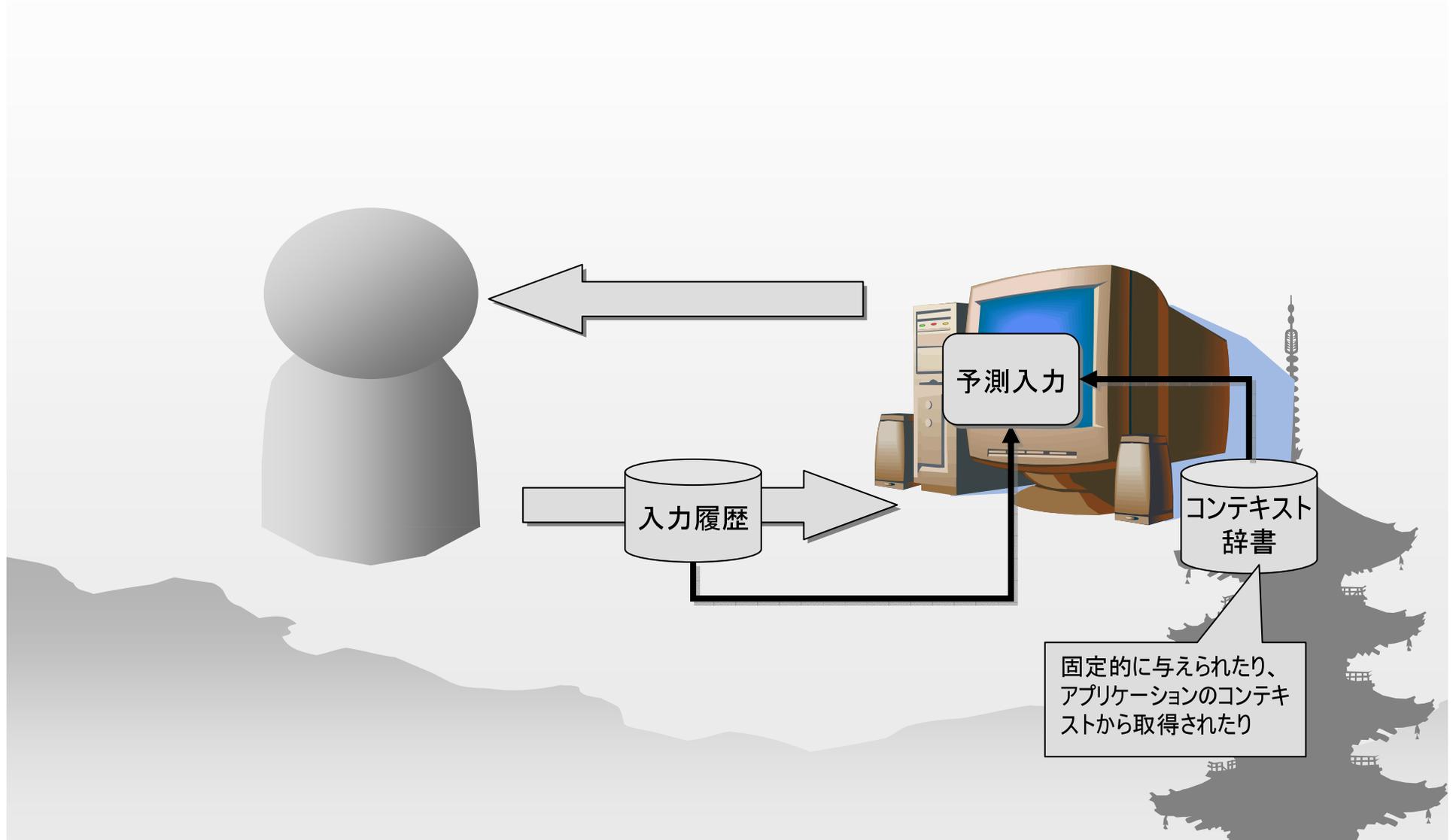


# 出力履歴を利用した予測入力

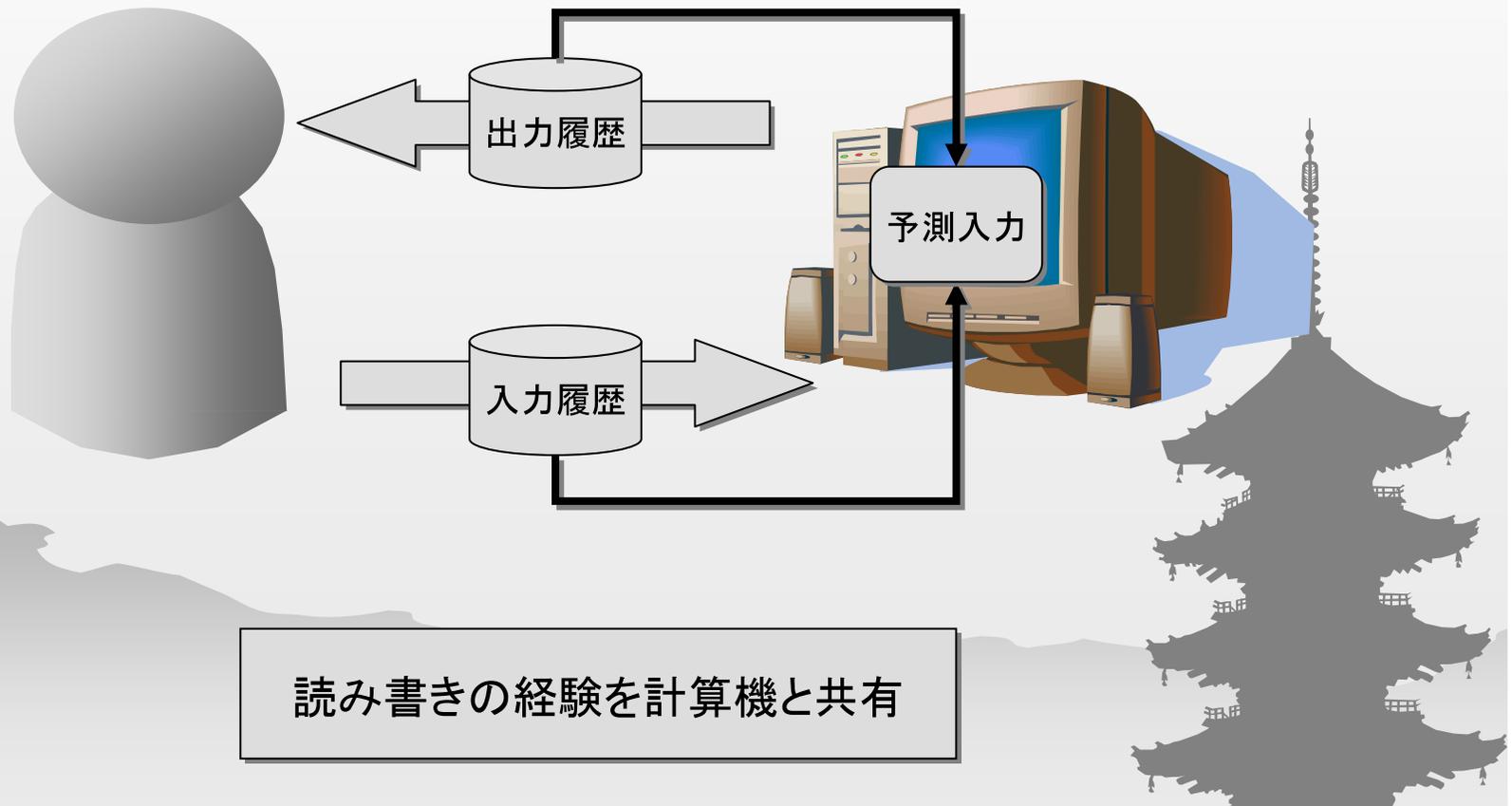
たなか ともひさ



# 従来の予測入力

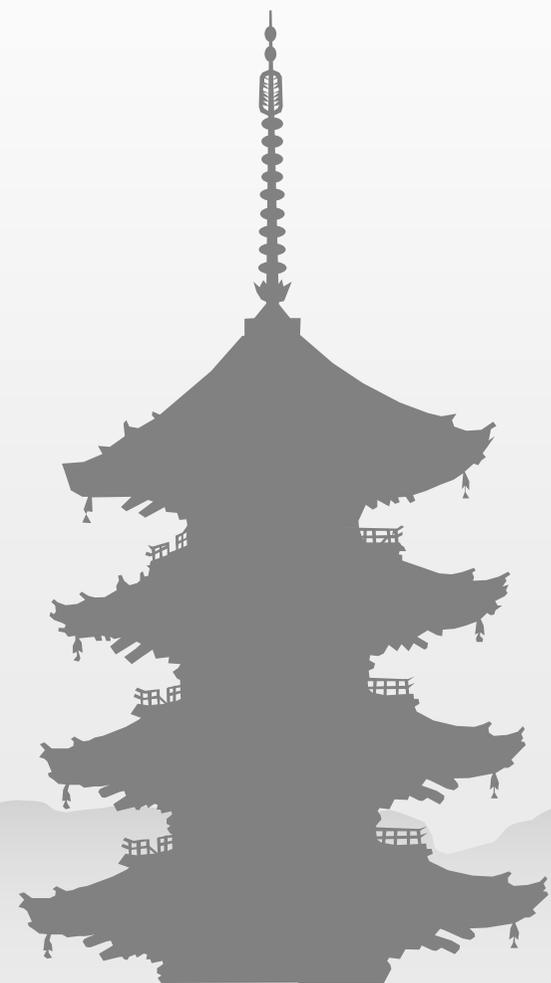


# 出力履歴と予測入力

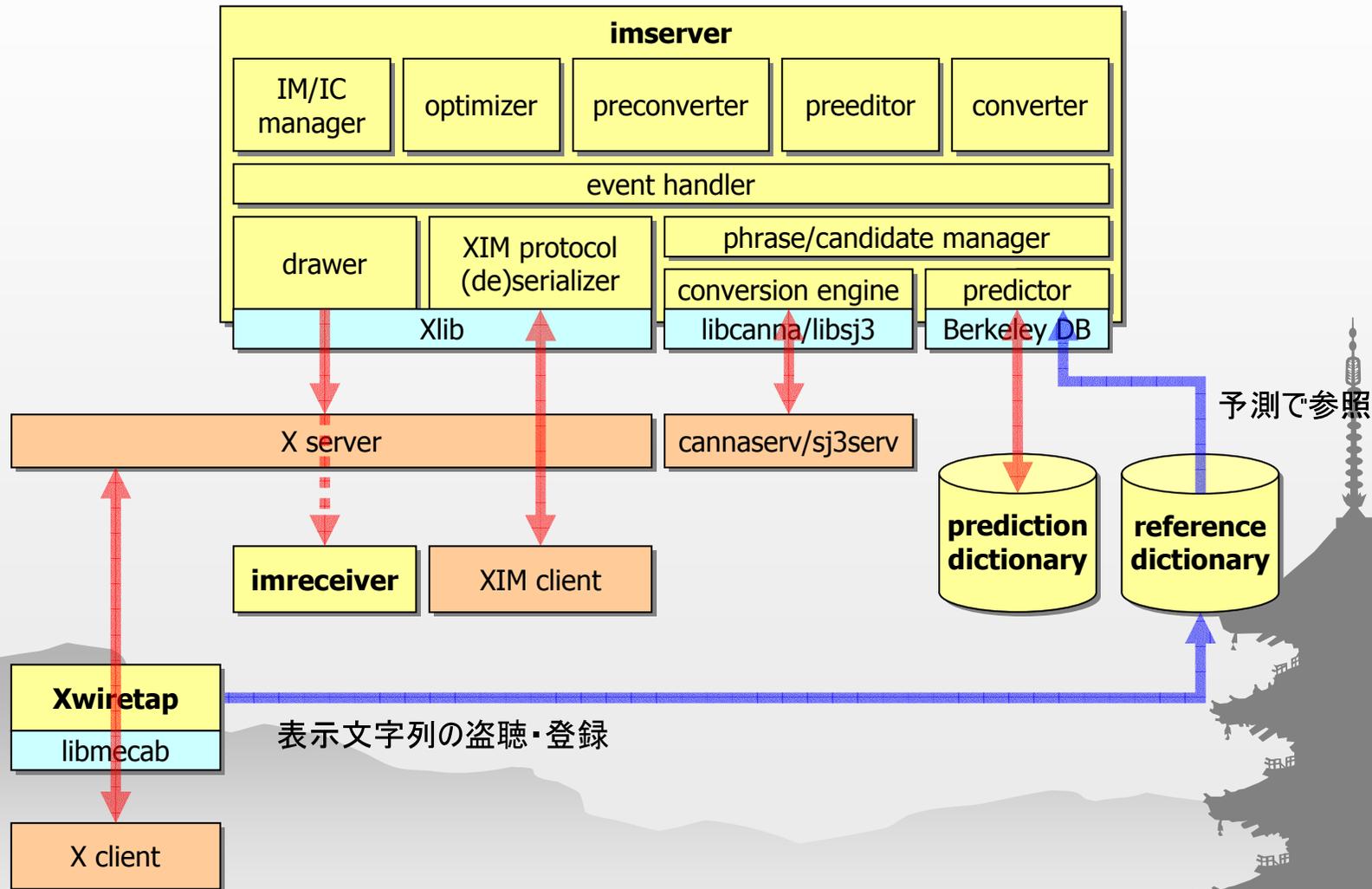


# 試作B

2004/03～2004/05



# 日本語XIMサーバ 試作B 構成

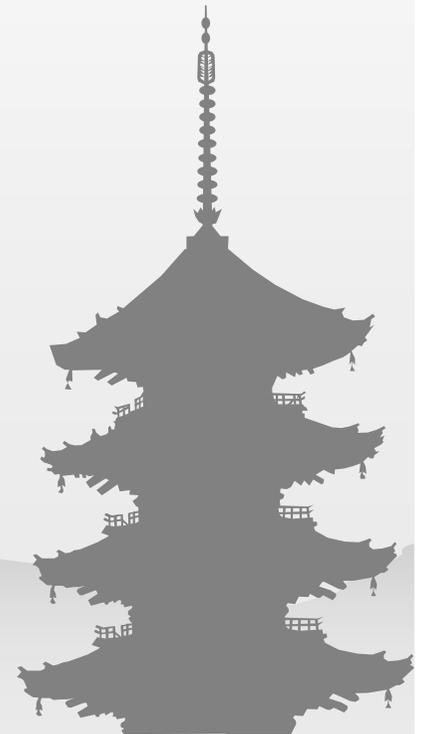


# Xwiretap

- ❁ Xプロトコルを盗聴するプロキシ
  - ディスプレイ番号#1への接続を#0に転送
- ❁ 描画リクエストに含まれる断片化された文字列から表示内容を推測
- ❁ 表示内容を文に分解
- ❁ 各文をlibmecabで文節に分解、読みを取得
- ❁ 文の先頭から文節を取り除きながら参照辞書に登録

# 表示内容の推測

- ❁ 文字列描画リクエストに含まれる文字列は断片化されている
  - 行の折り返し
  - 文字の色、フォントの変更
- ❁ ヒューリスティックで表示内容を推測



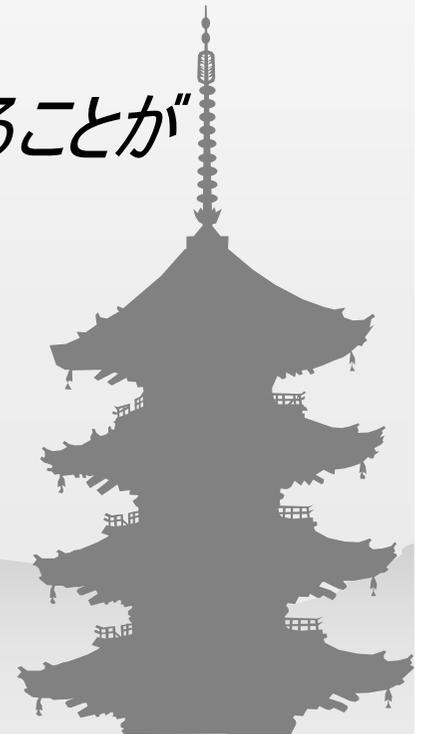
# 文に分解

## ❁ 推測した表示内容

- 「松嶋菜々子が今年6月に出産予定であることが、本紙の取材で明らかになった。」

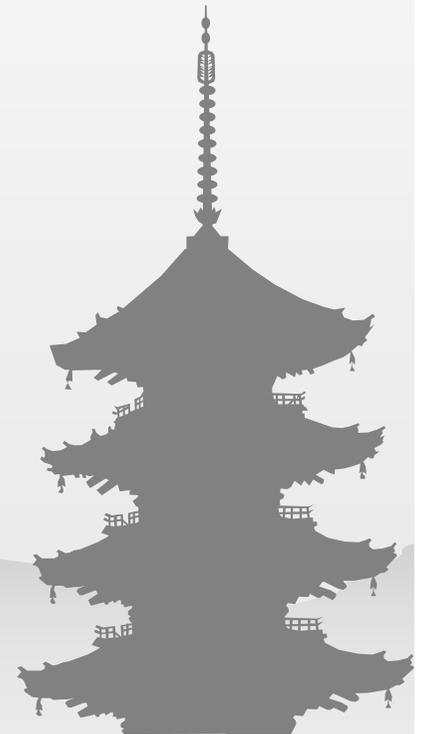
## ❁ 文に分解

- 松嶋菜々子が今年6月に出産予定であることが
- 本紙の取材で明らかになった



# 文節に分解（形態素解析）

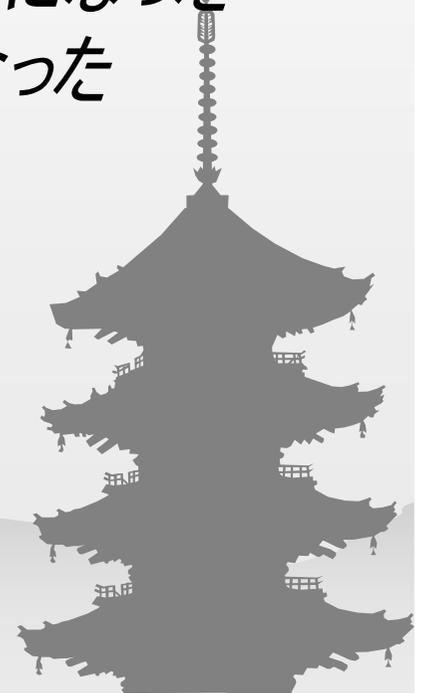
- ❁ 文をlibmecabで文節に分解
  - 本紙の取材で明らかになった
- ❁ 同時に各文節の読みを取得
  - ほんしのしゅざいであきらかになった



# 参照辞書への登録

## ❁ 文の先頭から文節を取り除きながら登録

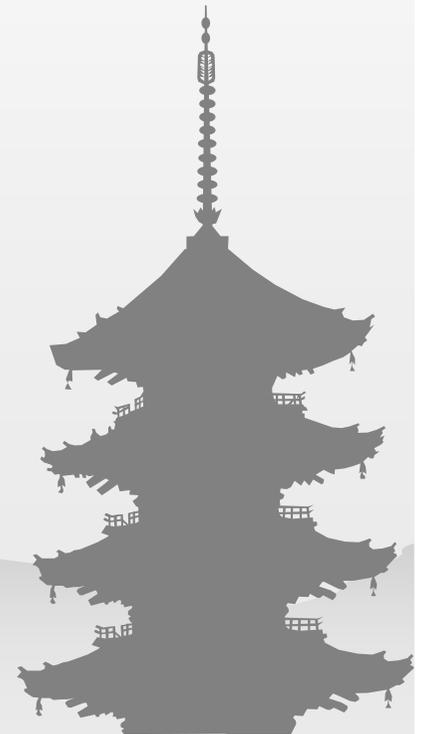
- ほんしのしゅざいであきらかになった:本紙の取材で明らかになった
- のしゅざいであきらかになった:の取材で明らかになった
- しゅざいであきらかになった:取材で明らかになった
- であきらかになった:で明らかになった
- あきらかになった:明らかになった
- になった:になった
- なった:なった



# 参照辞書への登録

## ❁ 文節を登録

- ほんし:本紙
- の:の
- しゅざい:取材
- で:で
- あきらか:明らか
- に:に
- なった:なった



# 予測候補の優先度

- ❁ 辞書への登録時の時刻が新しい順に表示
  - 最近入出力したものが優先
  - 入力履歴と出力履歴を同等に扱う
    - ➡ 参照したい文書を画面に表示するだけで、その内容が予測候補になる
- ❁ 参照辞書の利用の選択
  - 必要なときだけ出力履歴を利用できる

