大学共同利用機関法人 人間文化研究機構
国立国語研究所
National Institute for Japanese Language and Linguistics

言語資源研究系
萌芽・発掘型共同研究プロジェクト

統計と機械学習による日本語史研究

リーダー：　准教授　小木曽智信
〔共同研究員：6名〕

歴史的な
日本語資料の
テキストデータ

機械学習によるテキスト処理
濁点，句点の自動付与
仮名遣い，送り仮名の整備

UniDicによる形態素解析（短単位）
中古和文UniDic・近代文語UniDic …

長単位・文節への組み上げ

開発した技術を通時コーパス構築に活用

Oxford
VSARPJ Corpus

構文情報の
アノテーション

**単語解析済み
日本語史資料
データベース**

統計的手法を用いた日本語史研究に活用
語彙・文法・文体