

『日本語歴史コーパス』への追加情報の付与と共有 —中古和文の「る」「らる」を例に—  
小木曾智信

『日本語歴史コーパス』では、すべての語にその位置を示すユニークな ID が付与されており、これを利用することで用例を指し示すことができる。たとえば、『土佐日記』の中で最初に現れる助動詞「る」の用例（「守の館より、呼びに文持て来たんなり。呼ば【れ】て到りて、日一日、夜一夜、とかく遊ぶやうにて明けにけり。」）は、サンプル ID が「20-土佐 0934\_00001」、開始位置（サンプル先頭からの文字位置）が「5020」の語であり、「20-土佐 0934\_00001, 5020」という位置情報によって「中納言」の位置検索機能で参照できる。また、下記のような URL とすることで用例を指すパーマリンクとしても利用できるようになっている。

[https://chunagon.ninjal.ac.jp/chj/permalink?unit=short&position=20-土佐 0934\\_00001, 5020](https://chunagon.ninjal.ac.jp/chj/permalink?unit=short&position=20-土佐 0934_00001, 5020)

（position 以降が位置情報。「中納言」ユーザーであればログイン後にこのアドレスで当該例を表示・確認できる。）

この位置情報を用いることで、用例集を作成することができるほか、用例に用法分類などの研究で必要となる情報を付与して残すことも可能である。こうしたデータはユーザーによる追加情報の付与はコーパスへのアノテーションとして位置づけられ、コーパス本体とは切り離して配布し、研究者間で相互に参照、共有したり、再利用したりすることができるデータとなる。

こうした試みの例として、『日本語歴史コーパス』の中古の仮名文学作品（『竹取物語』『伊勢物語』『土佐日記』『紫式部日記』『更級日記』等）に現れる助動詞「る」「らる」について受身・尊敬・自発・可能の用法分類を実際に行い、コーパスの位置情報を用いて用法分類つき用例集を作成して下記のようにオープンデータとして公開した。

小木曾(2019)「CHJ 中古「る」「らる」用法分類アノテーションデータ」<https://researchmap.jp/mu1dor8so-12361/> 発表では、こうしたアノテーションデータを作成する方法と留意点、配布されたデータの利用方法について論ずる。