

テキスト平易化コーパスの構築指針

首都大学東京

梶原智之

小町守

kajiwara@jnlp.org

テキスト平易化

- = 語彙の平易化
- + 構文の平易化
- + 説明文生成

難解文

このシステムは、市街地での渋滞の主な要因となっている、交通量が多い交差点での信号待ちの車を減らす目的で、東京大学大学院情報学環の上條俊介准教授の研究室と警視庁が共同で開発しました。

文分割

日本語学習者のために...

- どのような種類の構文の平易化が必要か？
- どのような種類の語彙は換言で平易化できるか？
- どのような種類の語彙は説明文生成が必要か？

書下文1

このシステムは、市街地での渋滞の主な要因となっている、交通量が多い交差点での信号待ちの車を減らすことが目的です。東京大学大学院情報学環の上條俊介准教授の研究室と警視庁が共同で開発しました。

目的語の補完

書下文2

このシステムは、市街地での渋滞の主な要因となっている、交通量が多い交差点での信号待ちの車を減らすことが目的です。東京大学大学院情報学環の上條俊介准教授の研究室と警視庁が共同でシステムを開発しました。

節の挿入

書下文3

このシステムは、市街地での渋滞の主な要因となっている、交通量が多い交差点での信号待ちの車を減らすことが目的です。東京大学大学院情報学環の上條俊介准教授の研究室と警視庁が共同で信号待ちの車を減らすシステムを開発しました。

主題の交替

このコーパスがあれば...

- 日本語学習者のためのテキスト平易化のパタンの全貌が明らかになる
- 2010年～、英語のテキスト平易化タスクで主流となっているSMTによるアプローチも実現できる
- テキスト平易化の自動評価尺度も開発できるかも

書下文4

交通量が多い交差点での信号待ちの車は、市街地での渋滞の主な要因となっています。東京大学大学院情報学環の上條俊介准教授の研究室と警視庁が共同で信号待ちの車を減らすシステムを開発しました。

...

書下文24

交差点で信号を待つ車が増えると渋滞（＝道路が混んで、車が進まないこと）の原因の1つになります。東京大学大学院の上條俊介さんのグループと警視庁は、信号で待つ車を減らすためのシステムを作りました。

語の挿入

今後の課題

- 基本的な変換の連鎖への書き下しと変換ラベルの付与・分類を繰り返す
- 書き下しをクラウドソーシングで実現するためのアノテーションマニュアル
- 日本語教師に平易文への書き換えを依頼

平易文

交差点で信号を待つ車が増えると渋滞（＝道路が混んで、車が進まないこと）の原因の1つになります。東京大学大学院の上條俊介さんのグループと警視庁は、信号で待つ車を減らすための新しいシステムを作りました。