

1. SJ-CAT (Speaking Japanese Computerized Test)の紹介

SJ-CAT は日本語学習者のための日本語スピーキング能力をコンピュータ上で自動採点するテストシステムです。J-CAT® (Japanese Computerized Adaptive Test) のプログラムを基に作られています。テストは、文読み上げ問題、選択肢読み上げ問題、文生成問題、自由回答問題の4形式で構成されます。文読み上げ問題では、画面上に文が提示され、それを読む音声聞こえます。その音声の後に続いてその文を読み上げます。選択肢読み上げ問題では、質問に対して画面上に示された3つの文から適切な文を選んで読み上げます。文生成問題は、質問に対して一文で解答する形式です。自由回答は質問に対して、自分の意見を言うもので、45秒以内で話します。これらの音声を自動的に認識して、アクセント、イントネーション、流暢さなどの、主に音響的な特徴を基に点数化します。SJ-CATは人(日本語教師)による、話し方及び意味などの内容に基づく採点と同じような結果を出すことが確認されています。このセミナーではどのようにして機械が自動採点するのかについてお話しします。

1. Introduction to SJ-CAT (Speaking Japanese Computerized Test)

SJ-CAT is a system to assess Japanese speaking proficiency automatically on a computer. It is built on the basis of the system of J-CAT (Japanese Computerized Adaptive Test). SJ-CAT consists of four types of Q & A, reading a sentence, multiple-choice reading, generating a sentence, and open-ended answers. In “reading a sentence”, a test taker reads a sentence after a sentence is presented on the screen with its model reading sound. In “multiple-choice reading”, a test taker reads the best answer chosen from three sentences shown on the screen. In “generating a sentence”, h/she answers the questions by saying a sentence of her/his own. S/he is asked to state an opinion and give reasons for that opinion within 45 seconds in “open-ended answer”. SJ-CAT calculates scores by checking mainly acoustic features such as accent, intonation, and fluency. It is attested that the scores given by SJ-CAT correlates with scores given by human raters, Japanese language teachers, who scored ways of speaking and semantic contents of answers. In this seminar I will discuss how the machine assesses speaking proficiency automatically.

2. NLT (Ninjal-Lwp for Tsukuba Web Corpus)の使い方

NLT は、日本語のウェブサイトから収集して構築した約 11 億語のコーパス『筑波ウェブコーパス』(Tsukuba Web Corpus: TWC) を検索するためのツールです。国立国語研究所と Lago 言語研究所が共同開発したコーパス検索システム NINJAL-LWP (NINJAL-LagoWordProfiler) を利用しています。コーパスを検索する方法としては、KWIC(Key Word In Context)がありますが、これを使ってデータを分析するには専門的な知識が必要です。それに対して NLT はコロケーションの情報などを集計して示すレキシカルプロファイリングという方法で構築されていて、直感的に使うことができます。セミナーでは、教師と学習者が、この検索ツールで何ができるのか、コーパスという大量のデータから何が分かるのかについて、実際に NLT を使って実感していただきます。(セミナールームの WIFI に接続して使えるノートパソコンかタブレットをお持ちください。)

2. How to use NLT (Ninjal-Lwp for Tsukuba Web Corpus)

NLT is the search tool for Tsukuba Web Corpus containing 1.1 billion words collected from web sites. It utilizes the corpus search system NINJAL-LWP, jointly developed by the National Institute for Japanese Language and the Lago Institute of Language. KWIC is usually used to search words in a corpus; however, special knowledge is required to analyze data collected by KWIC. On the other hand, NLT is built on lexical profiling techniques to summarize information such as collocations, which enable us to use the tool intuitively. In this seminar, you, as a teacher or a learner, will experience what you can do with NLT and what you can find from large data in a corpus. (Please bring a laptop or a tablet which can be connected to WIFI in the seminar room.)