

IMAGACT-BR: Uma Ontologia da Ação do Português Brasileiro

O IMAGACT – Ontologia Interlinguística de Ação (www.imagact.it) é um projeto em realização na Universidade de Florença, Itália, que criou e desenvolveu uma infraestrutura online de linguagem baseada em corpora de fala espontânea de italiano, com 1.010 diferentes conceitos de ação representados visualmente por meio de cenas prototípicas, filmadas e animadas em 3D. Ao utilizar pequenos vídeos que descrevem ações básicas e apoiam a necessidade que o usuário/aprendiz tem de compreender plenamente a aplicabilidade dos verbos, especialmente em situação de aprendizagem de língua estrangeira, estimula-se, através de uma imagem mental, a descoberta do verbo que pode ser usado para descrever essa ação tanto na sua língua materna como na língua que está aprendendo e/ou analisando. Este processo envolve comparações semânticas e pragmáticas que ocorrem na mente do aprendiz ao considerar o movimento envolvido, o instrumento utilizado, a duração da ação, etc. O projeto IMAGACT foi criado a partir 1.200 verbos de ação encontrados nos corpora de fala espontânea do italiano C-ORAL-ROM (Corpora de Referência Integrado para Línguas Românicas Faladas - Italiano, Inglês, Francês e Espanhol - <http://lablita.dit.unifi.it/coralrom/>), e no English Corpus, retirado do British National Corpus. Na sequência, suas ocorrências foram identificadas e distinguidas e o resultado foi a criação de um banco de dados de conceitos de ação representados através de cenas em que o conjunto de atividades físicas mais frequentemente referidas no discurso oral das duas línguas, italiano e inglês, foi representado. Os verbos que compõem esse banco de dados podem expressar mais de um conceito e, da mesma forma, cada conceito pode se referir a mais de um verbo. Atualmente, o IMAGACT está disponível para consulta em italiano, inglês, espanhol e chinês. Contudo, o projeto continua em expansão e seu banco de dados está sendo incrementado com outras línguas como bengali, alemão, dinamarquês, hindi, magahi, manipuri, norueguês, odia, polonês, sérvio, sânscrito, tunisiano, urdu, além do português europeu e do português brasileiro. Nessa pesquisa, primeiramente, incrementamos esse banco de dados com o português brasileiro criando proposições que descrevessem cada uma das 1.010 cenas seguindo os mesmos critérios adotados pelo IMAGACT, ou seja descrevemos as ações apresentadas por meio de uma frase afirmativa, na voz ativa, utilizando apenas um verbo de ação no presente do indicativo. Na segunda etapa da pesquisa, estamos validando as proposições criadas através de metodologia baseada em competência. Na próxima etapa, iremos validá-las ainda de duas maneiras: 1- por meio de consulta a falantes nativos do português brasileiro para averiguar se

as proposições criadas para descrever as cenas em português são reconhecidas como conceituações adequadas, ou seja, se os falantes as reconhecem como construções típicas de sua língua materna e se correspondem a construções que usariam para descrever a cena em questão; 2- por meio de corpora do português brasileiro, prioritariamente, ainda que não exclusivamente, através do C-ORAL-BRASL (Corpus de Referência do Português Brasileiro - <http://www.c-oral-brasil.org>), criado e desenvolvido pelo Laboratório de Estudos Empíricos e Experimentais da Linguagem (LEEL) da Faculdade de Letras da Universidade Federal de Minas Gerais. Após as validações citadas, essa pesquisa pretende, ainda, analisar e descrever como falantes nativos do português brasileiro conceituam as ações representadas por cenas prototípicas e, por fim, realizar uma pesquisa de cunho sociolinguístico para averiguar se há diferentes formas de conceituar as ações representadas, principalmente no tocante à quantidade de verbos utilizados para conceituar uma mesma ação e, também, na frequência de cada verbo. Inicialmente, acredita-se que variáveis como idade, nível educacional/acadêmico e nível social poderão influenciar nos resultados. Com isso, serão selecionados grupos de respondentes que correspondam a esses perfis para suportar ou refutar as hipóteses estabelecidas. Ao final da pesquisa, além de incrementar o banco de dados do projeto IMAGACT com os verbos de ação do português brasileiro, espera-se contribuir com a descrição da ontologia da ação de falantes nativos do português brasileiro, além de estabelecer, ainda que não de forma exaustiva, os verbos de ação que são pragmaticamente mais frequentes no uso formal e informal da língua.

Palavras-chave: Ontologia da ação; Verbos de ação; C-ORAL-ROM; C-ORAL-BRASIL; IMAGACT; Corpora.