Modeling Multilingual Grounded Language
Bill Dolan, Principal Researcher, Microsoft Research

Language allows us endlessly creative ways to express the same basic meaning, whether through monolingual paraphrasing or through bilingual translation. That expressive power, though, poses huge challenges for computational approaches to language understanding: how can we model the relationship between strings that are superficially dissimilar and yet "mean the same thing"? This problem is becoming particularly acute as software interfaces move toward simpler, more natural interactions that often crucially rely on linguistic input. An intelligent interface needs to be able to decide, for example, that the following utterances all describe approximately the same user intent:

Show me some formal dress shoes
I need some shoes for a job interview
Ik wil kleedschoenen
Търся официални обувки
我需要一些正式 口合穿的鞋子

In addition to the challenge of modeling such mono- and multi-lingual alternations on a broad scale, we also need to learn to ground language in the real world – in this case, to an inventory of shoe styles, for instance, and show a set of relevant images/records to the user. How can we reliably map natural language utterances – no matter how they are expressed – to appropriate changes in machine state? This talk will describe and modeling program aimed at capturing how different expressions of the same meaning – whether within one language or across multiple languages-- are grounded in the real world. In addition, we will describe and demonstrate the results of crowdsourcing experiments aimed at building multilingual datasets that are grounded in video segments, database objects, programming functions, and human movement.