To be presented at the 10th Meeting of the International Pragmatics Association, 2007

The Role of Gesture in Inviting Back-Channels in Arabic

Yaffa Al Bayyari and Nigel Ward, Computer Science Department, University of Texas at El Paso

April 13, 2007

To be a good listener requires showing active listening, and this done in part with back-channels. To back-channel appropriately the listener has to realize when back-channels are welcome, and in several languages there are prosodic cues produced by the speaker which indicate such times. Visual cues are also known to play an important role in turn-taking, which raises the question of whether back-channels can also be welcomed by visual cues from the speaker.

For this study we used the UTEP Iraqi Arabic corpus of face-to-face free-content dialogs to determine whether visual signals produced by the speaker co-occur with subsequent verbal back-channel production by the listener. To do so we randomly selected for analysis one-second segments taken from times when the speaker was talking, excluding regions with laughter or overlapped speech. Each segment was labeled by a person with no knowledge of Arabic or of our aims. Using the Anvil tool, she marked each segment for the presence or absence of clear hand movements, clear head nods, and clear eyebrow movements. We also automatically labeled segments for the presence or absence of prosodic cues for back-channels, specifically either of two prosodic features previously identified: a sharp pitch downslope or a pitch upturn (Ward & Al Bayyari, 2006). Finally we determined whether each segment was or was not followed by a back-channel by the other speaker, specifically whether a backchannel began within 500 milliseconds of the end of the segment.

The labeling was done in two sessions. 116 segments were labeled in the first and 138 in the second. In the second session the labeler was of course more experienced, and also better familiarized to the speakers, having spent about two minute watching the initial portion of each dialog before starting labeling.

Our hypothesis was that visual cues co-occurred with subsequent back-channel feedback. Although there was no such effect with the first session labels, there was with the second session labels: overall 39% of these segments containing visual cues were followed by back-channels, versus 20% of those not containing visual cues (significant by the chi-square test). The tendency was even stronger for segments containing prosodic cues: 87% of those also containing visual cues were followed by back-channels, versus 37% of those not also containing visual cues. Hand gestures were the most common, and the best cues to subsequent back-channels.

The unexpected difference between the first and second session labels can perhaps be attributed to learning on the part of our labeler. After experience she remarked that she had learned that the speakers in the corpus were very expressive but she remained unsure whether they were making “significant” gestures more or less frequently than in her native languages (English and Spanish). Statistically, however, by the second session she had learned to notice more gestures, and the gestures she noticed were more “significant” in the sense of better relating to listener back-channels. This suggests the need to study cultural differences in gesture perception.