35th Annual Linguistics Spring Colloquium
Saturday, April 21 2012

9:00 AM – 3:30 PM
Room GMCS-301
San Diego State University
[9:00 AM – 9:45 AM] – A light breakfast will be served for morning attendees and presenters.


One of the outstanding open problems in the fields of machine translation and natural language processing is anaphora resolution. Anaphoric expressions are expressions whose reference depends on another referential expression (called its antecedent) in the discourse. In languages like English, anaphoric expressions are generally overt; for example the pronoun himself, as in John shaved himself, is an overt anaphoric expression whose antecedent is John. On the other hand, in languages like Japanese, anaphoric expressions are often deleted in discourse, giving rise to what are called zero pronouns. The issue of Japanese zero pronoun resolution has been discussed extensively in syntactic, discourse and computational studies. There are two kinds of approaches to zero pronoun resolution: rule based approaches and statistically based approaches. In the current study, I pursue a statistical-based approach, using what is called a Maximum Entropy model to resolve zero subject pronouns. Preliminary results have uncovered a class of remote antecedents that cannot be reliably detected by the model, suggesting that a much larger amount of data or linguistic knowledge (rules) may be required to achieve a higher level of accuracy.


Web texts have undergone an interesting transformation in the past decade. Humans increasingly interact with one another through social media. Depending on the site and the user’s audience, users deploy a vocabulary and illustrate their membership with other groups in ways that reinforce or transgress typical understandings of social identity. Because of these subtle shifts in the way users perform group membership, extant methods in user classification may fail to capture the peculiarities of human identity. This project will explore methods for user classification on a gay dating site. User classification refers to the task of sorting users based on the language encountered in natural language texts. Social networking and dating sites pose peculiar challenges for user classification. The kind of language encountered on these sites is conversational, colloquial, unedited, and informal. What’s more, social identity is oftentimes articulated in ways that on the surface seem contradictory. This project looks at one gay networking site where users have chosen a scene (leather, jock, trendy, to name a few) in the construction of their user profiles. A machine-learning algorithm was applied to the self-descriptions (performances) in these user profiles to automatically classify a user’s scene affiliation based on performance. The machine-learning algorithm returned better than chance results for scene-by-scene comparisons and when classifying all scenes. Because the meaning behind these kinds of human-produced texts is wrapped up in a symbolic space that is not easily tractable, this research looks at how social identity theory can be incorporated into improving the accuracy of future user classification tasks of this same nature.

Twitter’s current search functionality is Boolean in nature, thereby ignoring several foundational principles of current search philosophy and best practices. A brief introduction to search is given with particular attention being paid to the issues of polysemy and synonymy and their impact on query generation and parsing. A Twitter-specific use case for advanced search is given and implementation roadblocks of a generally better search feature for Twitter are discussed. Finally, a novel methodology which relies on well-known search techniques and off-the-shelf API methods to bypass said roadblocks is introduced, and a roadmap to finding a non-ranked set of tweets relevant to one or more of the latent topics associated with a query is laid out.


TimeML, TimeBank, and TTK (TARSQI Project) have been playing an important role in enhancement of IE, QA, and other NLP applications. TimeML is a specification language for events and temporal expressions in text. I will briefly introduce specification of English TimeML developed by ISO and Pustejovsky at Brandeis University and present the problems and solutions for porting TimeML to Korean as a part of the Korean TARSQI Project. I will also introduce the KTK which is an automatic markup tool of temporal and event-denoting expressions in Korean text.

[11:45 PM – 1:30 PM] — Break for lunch
[1:30 PM – 2:00 PM] – The Use of Wanna Contraction Among Native Korean Speakers of English (Sam Spevack, Undergraduate)

This study will investigate the ability of native Korean learners of English to show knowledge of when WANT and TO can contract to form WANNA. In English, while want-to contraction can occur in certain conditions (eg. Who do you wanna take to the beach?), it cannot in seemingly similar sentences (eg. *Who do you wanna take Amy to the beach?). WANT and TO are believed to be not able to contract when a question word (eg. who) originally intervenes between them before it moves to the front of the sentence. When acquiring a first language, children are hypothesized to rely on a set of rules, which are known as Universal Grammar (UG) and include the constraints on want-to contraction. It is debated, however, whether or not this same set of rules is able to be accessed when acquiring a second language after early childhood, or if second language acquisition is facilitated by the first language instead. In languages such as Korean, question words do not undergo movement from their original positions and constraints on want-to contraction are not relevant. Therefore, if native speakers of Korean exhibit access to the knowledge of the constraints on want-to contraction, this access must be facilitated by UG and not the first language. In a previous study (Kweon and Bley-Vroman, 2011), it was reported that native Korean speakers learning English in Korea did not show native-like access to constraints on want-to contraction. This study will examine the ability of native Korean speakers who have been exposed to English in the United States for over a year’s time. Participants in this study will include both native Korean learners of English and native speakers of English. They will be asked to give intuitions about the grammaticality of sentences with grammatical and ungrammatical uses of want-to contraction. If the native Korean speakers show native-like intuitions for sentences with want-to contraction, then this study will show evidence supporting access to UG during second language acquisition.

[2:00 PM – 2:30 PM] – Introducing the San Diego Sociolinguistic Documentation Project (Dr. Douglas Bigham, Faculty)

What do San Diegans think about the way they talk? Do they perceive differences between the way they talk and the way “other” Southern Californians talk? Does a San Diegan’s perception match up with reality—how do San Diegans actually talk? Using data collected as part of the new San Diego Sociolinguistic Documentation Project, in this talk we’ll explore the beginnings of these questions. Research participants were asked to perform a text with two puppets – Danny the Dolphin and Jaime the Bat. Comparing this performance to the participants’ interview speech provides insight into the salience of linguistic variation (Schilling-Estes 1998). When a performer “acts out” a dialect, he or she engages in an implicit discussion of perceptual dialectology, drawing on those dialect features which are most salient in order to construct the performance (Preston 1992). In an area undergoing sound shifts, these performances take on heightened importance, allowing a window into how new and traditional dialect forms can co-exist in the construction of a local identity (Bucholtz 2001, Trester 2008).
[2:30 PM – 3:00 PM] – IMPOLITENESS IN COMPUTER-MEDIATED COMMUNICATION (ERIC DALTON, GENERAL LINGUISTICS)

Among “encultured” net users, social niceties that, in past generations, seemed commonplace and nonnegotiable for maintaining close-knit communities and servicing solid interpersonal relationships appear to be all but dissolving as reputable virtues within computer-mediated communication (CMC). It is now disrespectful speech or impoliteness that often seems to govern the social interactions of many virtual communities. Two forms of impolite “speech” in particular, flaming—aggressive expressions of over-the-top antagonistic behavior—and trolling—the internet equivalent of skulking about an online community looking for trouble or baiting users into an argument—are most often exemplified by aggressive forms of affrontery-centric speech. As this phenomenon of online aggressiveness has become more pervasive and widespread, this thesis aims to take a closer look at the roles and pragmatic functions that antagonistic behavior plays in negotiating the social and cultural norms of an online community. In order to accomplish this, I apply a reconstituted form of politeness theory impoliteness theory, to two forms of online communication that represent distinctly different methods of online interaction. The resulting transcripts include texts from online public discussions, such as MSN.com, and recorded voice Chat in pre and post-game lobbies found within the competitive online multiplayer portion of the videogame “Halo 3.” Within the analysis to follow, this thesis showcases various unique idiosyncrasies of speech patterns exclusive to CMC discourses and further elaborates on how these net-centric facets of modern digital communication influence the interpretation of (im)politeness speech acts. Furthermore, this thesis investigates to what degree the established conventions of these CMC discourse patterns act as identity markers which help encultured net users to further establish identity and communicative solidarity within discourse communities. To summarize, this thesis concludes that speech acts of (im)politeness in online discourses do play a role in negotiating cultural/community norms. I further conclude that instances of flaming or trolling are not simply unmotivated acts of aggression, but are methods of expressing solidarity with others of particular ideological stances, means of establishing a makeshift pecking order between interactants, indicators of discourse community membership based on levels of enculturated discourse competence, or are similarly used for fulfilling the personal “face wants” for one’s own self-gratification at the expense of others.

[3:00 PM – 3:30 PM] – PERFORMANCE, HEARER-GIVENNESS, AND AUDIENCE ENGAGEMENT: A CASE STUDY IN KASHMIRI ORAL NARRATIVE (ASHA TICKOO, XI’AN JIAOTONG-LIVERPOOL UNIVERSITY)

Bauman (1975, 1977, 1986, 1989, 2000, 2004) suggests that performance is a specific kind of speech in which a very particular, formally coded kind of presentation ‘keys’ the oral literature/verbal art frame. I want to show that performance-keying mechanisms in “Hatim’s Tales”, a collection of oral narratives recorded and transcribed in the 1920s (Stein, 1923), effect a kind of collaborative engagement from a non-vocal audience. While Hatim’s Tales are widely acknowledged as making a significant contribution to the considerable body of Kashmiri oral literature, they have remained largely unstudied by scholars with interest in linguistic approaches to the assessment of narrative. I will suggest that they provide compelling evidence of oral literature as a genre that is essentially audience directed and dialogic, despite its ostensibly monologic character. Performance keying-mechanisms in Hatim’s Tales are put in place in various different ways, all of which draw attention to, and strategically utilize, information the hearer already knows (on given-new cf. Chafe 1976; Clark & Haviland 1977; Halliday 1967; Kuno 1972, 1974, 1978, 1979; Halliday and Hasan 1976; Horn 1978; Prince 1979, 1981, 1992; amongst others). I will first describe three distinct ways in which hearer-given information is used to key different kinds of performance genres, by putting in place distinct kinds of audience appeal. I will then suggest that this is part of a more robust performance-keying systematicity which uses hearer-given information to enhance audience-involvement in the developing narrative. The study aims to support the claim that keying mechanisms not only signal the performance of oral art, but also evidence the unique audience-directed and essentially communicative character of that art.
To all of our presenters and attendees; the faculty and staff of the Department of Linguistics and Middle Eastern/Asian Languages including Lois Marsico; our retiring faculty—Drs. Soonja Choi, Robert Underhill, and Zev Bar-Lev; our LSA Supervising faculty, Dr. Robert Malouf; our 2011-12 LSA officers—Jessica Campbell (President), Nick Carleton (Vice-President), Paul Tarpey (Treasurer), Sara Kazemi (Secretary), and Sam Spevack (Undergraduate Liaison); and our SDSU-affiliated funders—Associated Students Cultural Arts & Special Events (CASE), Instructionally Related Activities (IRA), and the College of Arts & Letters (CAL) council.