Blogs are the Mirrors to Ourselves: Examining Emerging Semantic Conventions in an Online World

Gretchen Trupp
December 23, 2017

Acknowledgements

I would like to thank my thesis advisors, Peter Klecha and Jonathan Washington, for their guidance and direction throughout this process, and Jeremy Fahringer for technological support.

1 Introduction

This paper will examine several emerging online conventions and their functions in English and Chinese internet varieties. The internet has created an environment that encourages and sometimes necessitates the creation of new linguistic conventions in text-based communication. Emerging semantic conventions in written language varieties have entered the internet lexicon seemingly to remedy the issue that some conventions, such as prosody, are lost in the conversion between spoken language and internet language. The internet as a communication medium is unique to its written predecessors in that it is instant and archived mass communication that is accessible to billions of interlocutors. It creates an imperative for speakers to develop the tools to parse ambiguities that arise from factors such as an inability to read facial expression, prosody, etc. (Baron [2003]). Where speakers feel a need to reduce ambiguities of this nature, internet language conventions emerge.

These phenomena are not unique to widely used languages on the internet such as English in that similar conventions are emerging among internet users.
worldwide in multiple different languages. Interestingly, the conventions also convey natural meanings, such as the emotional state of the speaker and speaker identity. Here, natural meaning refers to factors conveyed by an utterance that are not necessarily semantically or pragmatically related to the content of the utterance, but are socially constructed inferences that can be drawn about the speaker's emotional state or identity (Grice [1957]). These features are usually deducible in spoken or signed languages, but standard written conventions for reproducing variables such as prosody have previously not existed. Using new conventions in a variety of contexts, speakers are able to more effectively communicate their identities in a digital space. These emerging linguistic features work to create a space for communication that fluidly transcends geographic location, and thus facilitates widespread identity formation through consumption and replication of linguistic signs posted on the internet.

The internet, as a semiotic space (Blommaert [2013]), generates meaning for interlocutors in an unprecedented fashion, in that a digital platform allows large amounts of information to be created and consumed, and for interactions to take place regardless of physical location and time. These interactions, however, can lack factors such as facial and manual gestures and prosody, which can be critical to successful communication between participants. These communications also have the ability to fluidly move between synchronous and asynchronous modes of dialogue (Crystal [2001]). Establishing new conventions facilitates communication that can transcend the barriers inherent in a textual medium using these new communicative modes. Studying these emerging linguistic phenomena is important for understanding how language and identity intersect, especially in newer online communities that have not received much scholarship. The goal of this paper is to examine a select few new conventions, their function in the online lexicon, and how these new online conventions remedy semantic problems unique to a textual medium in this new global digitized landscape. This is not an attempt to exclusively categorize and describe new trends, but describe how these trends function in a global digital space at this point in time. The conventions that will be examined are drawn primarily from internet English, with some analysis of internet Chinese conventions as well. The English conventions that I examine are the circumfix “/” around a word of phrase, the clitic TM, and suprasegmental spacing between letters in words. The Chinese conventions that I examine are the sentence-final modal particle “/” and the use of pinyin to avoid marking gender in the speaker. Even though
this is not a huge part of my analysis, I will also be looking at numeric usage trends on Twitter for 2013-2017. In order to analyze these trends, I draw on semantic utterance observation schema as well as sociolinguistic methods of discourse analysis in media and semiotic spaces as developed by Blommaert [2013]. This synthesis of analytic methods ensures that I have described the conventions as accurately as possible, and that my analysis takes into account the spaces these conversations take place in and the interlocutors themselves. These conventions were examined using a Twitter corpus found on archive.org as well as random searches of Weibo. A further explanation of methods can be found in Section 2 below.

2 Methodology

Due to the differing availability of robust corpora for data acquisition, the methods and thus the analysis are different for the English and Chinese data that I have collected. The English data is collected from the Twitter Stream API by the Archive Team at archive.org called “The Twitter Stream Grab”, processed to include only English-language tweets, and then was exported from JSON data format to tab-delimited subset of data fields for each tweet. The Chinese data is drawn from random searches of Weibo microblogs. Both websites are popular microblogging spaces in the geographic regions where these languages are most commonly spoken, and are similar enough in form and function to serve as comparable to each other in terms of analysis (Chen et al. [2011]). Microblogs are accessible to a wide range of users that access microblogs actively, and “[are] in particular marked by innovative and varied use of orthography”, which is important in observing emerging conventions (Ling et al. [2016], 2). Due to the lack of prepared Weibo corpuses, my data and therefore my analysis are not as complete for my Chinese data as for my English data. Instead, my Chinese data enriches my analysis of written English conventions as well as demonstrates the global need and reach of emerging internet conventions. In order to most effectively describe these conventions, I also acquired informal judgements from native speakers to further analyze these online utterances. With raw data acquired, I analyze both specific instances of the convention and its larger role in internet discourse. My Twitter data was collected using the following regular expressions using an example data file containing a month’s worth of tweets from June 2015, as well as the regular expression for the total number of tweets each month
that were gathered in the Twitter Data Stream:

Total Tweets in the Collected Data Stream:

time bzgrep -c "\ten\t" tweets_2015-06*

Circumfix “~/”:

bzgrep "~/\w.*\w~" tweets_2013-06_en.dat.bz2 | grep -v "RT "

Clitic TM:

bzgrep "TM" tweets_2015-06_en.dat.bz2 | grep -v "RT "

Suprasegmental Spacing:

bzgrep "[a-zA-z][a-zA-z][a-zA-z]" tweets_2015-06_en.dat.bz2
| grep -v "RT "

It is important to note that these examples are not necessarily generalizable to the uses of these conventions across the entire Twitter community, because the corpus used collected the lowest stream of data, producing monthly results in the thousands of these phenomena—when in reality observing data from the entirety of Twitter archives would produce much greater results. Some pitfalls of this analysis are that the corpus only captures data from 2013-2017, and so it is hard to definitively draw conclusions on the exact emergence of the Twitter data or make claims regarding all Twitter activity during the examined periods of time. Because there is no corpus for the Chinese data, this data is only generalizable to those random searches and not necessarily reflecting the full scope of the convention. In order to protect the anonymity of the users and the validity of the data itself, all data containing usernames or identifying information has been omitted. At the time this paper was written all posts were publicly available. In the future, observing these conventions over a longer period of time as well as acquiring a more robust corpus for the Chinese data would allow for greater generalizability.
3 Background

If human writing has existed for thousands of years, why haven't written conventions that serve the same function as prosody emerged prior to the internet? While each language community is different, various social, technological, and spatial factors have prevented these conventions from developing a widespread and lasting reach until now. Previous communication, both spoken and written, has not had the global reach that the internet currently does. Written communications have historically been either private one-on-one communication, or taken the form of a sign, book, or some other document meant to be observed or consumed by interlocutors in that specific space. These signs are still very much a part of the internet today, but a variety of other communications are also taking place in this shared space. Utterances themselves can be shared between platforms, taken from one-on-one communication to public discourse, and disseminated to millions of people in a matter of seconds. “The reach of the internet is global, perhaps more extensive than any other communication medium” (Holt [2004], 79). With so many users all using one medium to communicate from a variety of locations, it is much easier to create a new convention in a medium that most of the speakers of the target language have access to. There are fewer powerful institutions on the internet that can perpetuate certain prescriptive conventions by enforcement, as opposed to other written forms of language which tend to also coincide with a more formal register as opposed to the generally less formal register of internet communication (Baron [2005]). The more informal register of the internet makes it much easier for language change to occur in a written medium, as that medium is accessed by an incredibly large user base relatively free of the prescriptive enforcement that has historically tended to come with more formal written language. While written communication on the internet bears obvious resemblance to more formal written communication conventions, additional conventions that mimic spoken language have also emerged.

Internet communication, computer-mediated communication (CMC), and netspeak are all ways to refer to discourse that takes place online (Baron [2003]). Mentioning aspects of one’s identity, using emoticons, and other types of language use are all methods that individuals use to communicate and explore their identities in an online world, creating an online persona for the public to see (Huffaker and Calvert [2005]). Internet communication has grown from sending strings of data between computers in the 60’s
with the first emergence of the World Wide Web in the 1990’s. This was a pivotal step for both the continuance of this medium as a form of communication and as a medium that did not require much outside training to use. In less than thirty years since the emergence of the World Wide Web, both synchronous and asynchronous communication have developed rapidly in the form of IM, blogs, and social media (Baron [2008]). On the internet, synchronous communication consists of two people communicating at the same time, such as instant messaging apps or chat rooms where two or more individuals are talking to each other in real time. While this has the potential to be asynchronous as well, its main communicative purpose renders it mostly synchronous. Asynchronous communication, on the other hand, is communication that does not have to happen within one defined time sequence, such as replying to a Facebook post or forum thread. Now, speakers have increased ability to communicate with an interlocutor synchronously despite time or physical differences, or have the interlocutor be able to interact with the utterance at a different time, with an additional caveat: this was a new medium, which came without an intense written prescriptive presence. This was also different from previous written communication mediums in that utterances are present indefinitely after they are uttered in some form of a catalogue— they will be forever stored and accessible unless actively deleted. It was much harder to communicate synchronously and asynchronously with an interlocutor when the physical distance between the speakers was great. While letters were an earlier form of asynchronous communication, they were not preserved in the way IM, email messages, or blog posts are today. The internet removed preservation and distance barriers, and with its development came a greater capacity for interaction with an interlocutor despite barriers that would prevent spoken communication or would have forced prescriptive constraints on the speaker. The internet is also more accessible than previous mediums as one does not need “expensive equipment or exhaustive technical training to compose effective utterances”, unlike the composition of written letters or a use of a printing press (Holt [2004], 131).

Additionally, the internet is a space that allows users to perform their identities in various ways, such as adhering to conventions or norms imposed by a group occupying a specific site on the internet, using specific lexical items to signify in-group attachment, and posting specific content (Crystal [2001]). One such example is that when users type, many will break their written utterances into smaller pieces, dividing their “turn” in the discourse, which is known as utterance chunking (Baron [2008], Ch.4). A few examples
of utterance chunking is shown below, taken from my personal online internet conversation about purchasing snacks for an event at two grocery locations—Giant, and the local Co-op:

(1) a. X/Y: Snacks
b. X/Y: Chips and salsa
c. X/Y: Lemonade
d. Y/X: Decided not to go to giant
e. Y/X: Will go to the co op tho

(2) a. X/Y: So when are we getting food
b. Y/X: Well i think im still gonna go after to class
c. Y/X: To the barbershop
d. Y/X: Ill see
e. Y/X: My phone is almost dead tho
f. X/Y: Major rip

Here, both users X and Y employ utterance chunking in their messages. One usage is speaker X making a list in the first example, while the usage in the second example is Y clarifying potential ambiguities in their first part of their utterance followed by breaking up their decision-making process. As Baron [2008] notes, utterance chunking can take place in variety of positions, usually with conjunctions. These conventions are much more similar to face-to-face speech as opposed to conventional writing, but common internet speech phenomena such as abbreviations and acronyms are not (Baron [2008]). The internet has created a space for a more informal register than what conventional writing is typically used for, such as formal papers or personal letters (although this register is certainly still present on the internet and can be informal to some extent, even when not online).

The internet as a new medium also allows for users to engage with their beliefs in a variety of new ways. Users can use search engines to seek out information, find groups of people that think similarly, utter their political beliefs and interact with millions of others, as well as actively engage in discourse with those who think differently than them. Social media, particularly in countries such as China where people, particularly students, are generally considered by the national and international public to be apolitical (Kim [2008]), “lowers the threshold for people to engage in online discussions, cam-
campaigns, and mobilization... [it also] enable[s] people to find a voice and share concerns that before were hidden from the public view” (Svensson [2016], 69). This interaction with the public is the key factor in setting the internet apart from other written mediums of the past: the ability of the public (that is, anyone with access to the internet) to read and engage with utterances produces an environment that is not only widespread with a large user base, but is incredibly varied in content. The internet has even been able to define what the public is— anyone who might be following someone, their friends, or potentially everyone who has access to the internet. This large user base with varied identities and opinions potentially sharing the same spaces can easily become a hotbed for disagreement and violation of cooperative principles and problems with written mediums emerge— but the internet also creates an environment where users can create new conventions to rectify these issues.

Because there are so many users from so many geographic locations, the internet operates as a multiracial, multiethnic, and multilingual space that can give historically oppressed voices a communicative platform. On the other hand, visual identity indicators are more difficult to deduce online, and so language use becomes the primary method for communicating aspects of one’s identity, and spaces that have historically been dominated by white Standard American English speakers are being relocalized as globalization occurs, such as Hawaiian language revitalization efforts using the internet (Warschauer [2001]). In order to combat potential ambiguities that arise, users have written in their own languages/dialects and created new conventions for expressing these identities. Some of these methods include communicating with their interlocutors salient identity aspects such as gender or race, and other conventions make the speakers’ emotional state less ambiguous for the interlocutor. The following English and Chinese conventions are used by a variety of internet language speakers to communicate these salient identity aspects and emotional states.

4 English Conventions

4.1 Clitic “TM”

Historical written use of the symbol “TM” has been exclusively related to businesses or corporations looking to trademark a symbol, logo, or slogan,
and still retains this meaning today ("International Trademark Association" [2015]). More recently, however, the use of "TM" has extended in use from just noting legal trademark status to new semantic meanings in different contexts. I argue that addition of this symbol/emoji carries metalinguistic meaning. All inferences conveyed by an utterance without "TM" are still conveyed when "TM" is inserted, and in addition use of "TM" can convey that whatever in the utterance is within the scope of the clitic should not be interpreted compositionally. "TM" can also convey that the host of the clitic is the stereotypical iteration and conception of the host, usually with highest prestige and/or standing in the eyes of the community. If it does not occupy the highest prestige, then it instead signifies that the host is being used within a particular discourse context for a particular community. For example, for internet users, debates on the internet can often be circular, remain unresolved, and involve a lot of ad hominem attacks. Thus, when the phrase "The Discourse™" is used, it refers to this discourse context for that community, which in this case is anyone familiar with that type of internet debate. This function is similar to the original non-compositional meaning that a trademark carries as a brand marker, just without legal backing. The two newer usages can create some ambiguity between meaning, especially when used for sarcastic purposes. This tweet below, gathered from the data, demonstrates the non-compositional meaning:

(3) a. A: BITCH IM CRYING IN THE CLUB™

In this tweet\(^1\) by user A, the user posted an image of the spot where Barack and Michelle Obama first kissed (See Appendix 3), with the caption "BITCH IM CRYING IN THE CLUB™". The phrase "crying in the club" is itself a reference to a snapchat that comedian Lil Duval uploaded of him crying captioned: "I don't know why I'm crying in the club right now". It then became a meme noting an emotional response to something that does not necessitate that level of emotion for that space. Through the use of the ™ symbol, user A references this meme as a way to communicate their emotional response to their posted image. Since this meme was so widespread, placing ™ after the phrase references that the speaker is in a highly emotional state in an inappropriate location. It can be argued that people who use Twitter and internet English often have either seen the meme or know the phrase, and putting the ™ after the phrase indexes the

\(^1\)Corresponding full tweets can be found in the Appendix.
meaning the original was trying to convey. Even without the knowledge of that meme, the interlocutor could still potentially understand the emotional response if they interpreted the “TM” to mean they should not interpret the utterance literally. Then, the phrase “crying in the club” would be just referring to a saddened emotional state in an unlikely place, because the tweet would still contain the necessary information relating to the emotional state of the speaker. “TM” conveys the non-compositionality of this phrase, and informs the interlocutor that the phrase is not to be taken literally. Referencing the meme is not necessary to convey emotions and mood to the interlocutor, but in doing so the tweeter performs their emotional state in a humorous redundant way using metalinguistic conventions.

(4) a. B: My heid is an erratic mess of ideas. Constant. And no just with food. Just constant fucking mental overload
   It’s why I like Twitter.
   b. C: @B that’s livin™

In the tweet from June 2015 shown above, user B first tweets about how their heid (Scottish English for “head”) is an “erratic mess of ideas. Constant...” to which user C replies, “that’s livin™”. Here, “livin” is the host of the clitic, and appears to be conveying the inference that the host, which in this case refers to B’s tweet about their mental state, as one that is the most stereotypical iteration of livin. This does not appear to be the state of living that carries the highest overt prestige, and so the interlocutor can potentially infer that user C is either being sarcastic or nihilistic, or saying that this is in fact the highest or most relevant standing that livin is capable of for a certain group of people that this user is a member of.

Both of these tweets demonstrate the non-compositional and stereotypical meanings that this convention has taken on. It also appears that the use of this convention could give rise to both implicatures, or just one at a time, depending on the surrounding content and the host itself. Parsing all of the potential nuances of this convention is beyond the scope of this paper, but it is clear that this convention has taken on these new meanings as it emerges as a more standard convention in the internet lexicon.

An initial search on Twitter for this convention in the month of June across several years brought thousands of results, not including retweets. The number of original tweets using “TM” during the month of June for the
following years is shown in the table below:

<table>
<thead>
<tr>
<th>Year</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clitic “TM” Raw Occurrences</td>
<td>2897</td>
<td>4296</td>
<td>3797</td>
<td>4676</td>
<td>4521</td>
</tr>
<tr>
<td>Total Tweets in the Data Stream</td>
<td>39462207</td>
<td>39766367</td>
<td>32690423</td>
<td>36519753</td>
<td>30380539</td>
</tr>
<tr>
<td>Percentage of Tweets</td>
<td>0.00734</td>
<td>0.01080</td>
<td>0.01162</td>
<td>0.01280</td>
<td>0.01488</td>
</tr>
</tbody>
</table>

Based on a few sample months of data, it can be deduced that this convention appears to be growing in popularity, which the raw number of original tweets almost doubling over the course of five years and the percentage increasing from approximately .007 percent to .01 percent. While some of these appearances are undoubtedly from companies or organizations using this convention for its more traditional brand meaning, the data shows an increased presence of the convention being used as these two clitics rather than just a legal trademark.

4.2 Euphemistic Circumfixation using “~ / ~”

While not as recently viral as “TM”, putting tildes as a circumfix around words and/or phrases is a popular way to perform the same function that a certain prosodic function performs in spoken English, namely euphemism or the sarcastic/self-aware attitude of the speaker. There has been some online discussion on the newer usage of the tilde (Bernstein [2015]), with some arguing that it is supposed to mimic vocal prosody and to mark what is between both tildes as linguistically significant, as well as convey speaker attitude such as sarcasm or snark (Leslie [2017]). Although these discussions are informal, they provide a framework with which to describe this phenomenon, and articulate that large communities of speakers find this usage grammatical. The tweet below is an example of one of the more recent usages of the “~ / ~” circumfix, where user D discusses their feelings about musical artists:

2 An important caveat is that this search contains all of the usages of “TM”; as a legal symbol, non-compositionality marker, and the stereotypicality marker.
(5) a. D: real shit tho Tyler the Creator is a rly important creator to me and so is Kanye and ik they’re both ~problematic~ so. don’t bother telling me

In this tweet, the word “problematic” is circumfixed by tildes. The user claims that this is an attribute that both Tyler the Creator and Kanye West share, but this does not change their opinion of them both as “important creator[s]”. Thus, the use of the tildes lets the interlocutor know that they are already aware of bad things the creators have done, and describe this with the euphemism *problematic*. It can be argued thus that the tildes here also mark sarcasm, since the user either does not think what the artists have done warrants the label *problematic*, or they just do not think that the label negatively affects their impact as artists. This is similar to the function and metalinguistic meaning of scare quotes, which is already a typed written convention, but is more academic in nature (Predelli [2026]). This discourse takes place on Twitter, which as a microblogging platform uses a more casual register. While scare quotes serve this similar function, tildes represent a more casual way to convey this prosody.

(6) a. E: @F i accidentally saw your sekai tweetfic and sobs loudly because too ~cuuute~ my mouth hurts from similing too much ;A;

In the tweet above, user E replies to user F about a *tweetfic*, or fanfiction story written on Twitter, that the latter user wrote that made them smile and cry because of how cute it was. Here the lengthening of the orthographic vowel *u* in *cute* as well as the tilde circumfix mimic the vocal prosody that would be used to intensify *cute*, communicating to the interlocutor that *cute* is incredibly linguistically significant— it is the cuteness of the tweetfic that has caused such an emotional response from user E. This emphasis, whether as an intensifier or as a sarcasm marker draws contrasts with alternatives, which signal an underlying convention rather than a purely cosmetic choice.

(7) a. G: are chris rock and “the rock” related?
   b. H: um, is that a real question
   c. G: @H im offended you thought i was serious
   d. H: @G It’s called.. wait for it...~*sarcasm*~
In this tweet from 2013, user G and user H are having a conversation about G’s joke, the first exchange in this talk. G jokingly asks if comedian Chris Rock and actor Dwayne “The Rock” Johnson are related. H then responds, seemingly unsure if G was asking a genuine question. G appears to take H’s response as a genuine question, and replies to let H know that they were offended that Chris Rock and “the rock” actually being related could be interpreted as anything other than a joke. H then replies that the original reply (“um, is that a real question”) was actually what was sarcastic, and was not meant to be taken literally. This is a clear case of H mocking G for not originally being able to deduce the emotional state of the reply, and so further augments the sarcastic tone by adding the tilde circumfix for linguistic significance. An interesting note is that in addition to “~/-”, [*] is also circumfixed. This also conveys the linguistic importance of what is being circumfixed, and has been previously documented as being an emphatic marker (Stede et al. [2006]).

From just the few examples shown, it becomes clear that the tilde marks various types of prosodic speaker attitudes such as sarcasm and snark. The squiggle shape of the tilde itself seems to also replicate the rise and fall of pitch that tends to accompany English prosodic sarcasm, making this a case of iconicity.

Below are the results of my initial searches for “~/-” as a circumfix appearing in the month of June across several years, which does not include retweets:

<table>
<thead>
<tr>
<th>Year</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>“~/-” Circumfix Raw Occurrences</td>
<td>9107</td>
<td>8005</td>
<td>6045</td>
<td>5448</td>
<td>4231</td>
</tr>
<tr>
<td>Total Tweets in the Data Stream</td>
<td>39462207</td>
<td>39766367</td>
<td>32690423</td>
<td>36519753</td>
<td>30380539</td>
</tr>
<tr>
<td>Percentage of Tweets</td>
<td>0.02308</td>
<td>0.02013</td>
<td>0.01849</td>
<td>0.01492</td>
<td>0.01393</td>
</tr>
</tbody>
</table>

2013 appears to be the height of popularity in this data set, as shown by the decrease in percentage of original tweets in the following years. This is consistent with the existing informal scholarship of this convention, which discusses examples from 2014 and earlier at the height of viral popularity.
4.3 Suprasegmental Spacing

As an internet phenomenon, one of the first instances of inserting spaces between the letters of words was used by the electronic cyber music group Vaporwave as an aesthetic touch to their album art. This group and in turn the typing style present on their albums grew in popularity in the mid-2000s ("Know You Meme" [2017]). This typing style involves taking a word and inserting spaces in between each of the letters to produce text that is spread further apart and thus takes up more physical space as text. The contexts and meaning in which suprasegmental spacing can be used has since expanded to a variety of other written discourse as a marking convention. There is a prosodic pattern or family thereof consisting of factors such as pitch, intensity, and lengthening that convey emphasis and emotion (Erickson et al. [1998]). Lengthened prosody, in particular, has already been documented as having an emphatic meaning (Herment-Dujardin and Hirst [2002]). Exaggerated prosody is a spoken convention in English varieties that has previously lacked a standard/ normative convention in text-based communication, and I posit that this suprasegmental spacing originally arose from this Vaporwave text style but over time changed and emerged as a convention to fill the communicative gap between written discourse and exaggerated prosody.

(8) a. I: Ok y'all lets get something straight Black Panther is a BLACK movie not a "POC" movie. It is a BLACK movie

In the tweet above, user I employs this convention in a status about the movie Black Panther. They are tweeting about how the up and coming Black Panther movie is a movie specifically for Black people, rather than all people of color. In order to emphasize their point, they use this suprasegmental spacing the last sentence of the tweet. The meaning of this suprasegmental spacing is made especially apparent in the contrast between "black" in the first and second sentences. In the first sentence, "black" is written with all capital letters, and in the second sentence to reiterate their point, they add suprasegmental spacing.

(9) a. J: My shift is dragging

Here, user J is describing how their shift is going extremely slowly. The length of their shift as well as their exasperation is made clear to the interlocutor with the use of suprasegmental spacing. Putting spaces between
dragging not only takes up more physical space when typing, mimicking the perceived length of the shift, but also mimics when speakers elongate a word in their speech for emphasis and exaggeration.

In both tweets, the one word suprasegmentally spaced communicates to the interlocutor that this word is marked as linguistically significant, and the lengthening of the letters, not only taking up more physical written space but also using up more characters. In a limited-character medium, each character must be carefully chosen in order to communicate the point, and so use of this convention appears to demonstrate a written form of lengthened prosody and its emphatic/exaggeration markers.

Below are initial searches of this convention used in tweets during the month of June across several years, which do not include retweets.

Table 3: Suprasegmental Spacing vs. Total Tweets in the Data Stream in June from 2013-2017:

<table>
<thead>
<tr>
<th>Year</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suprasegmental Spacing Raw Occurrences</td>
<td>90657</td>
<td>69820</td>
<td>46294</td>
<td>43320</td>
<td>33187</td>
</tr>
<tr>
<td>Total Tweets in the Data Stream</td>
<td>39462207</td>
<td>39766367</td>
<td>32690423</td>
<td>36519753</td>
<td>30380539</td>
</tr>
<tr>
<td>Percentage of Tweets</td>
<td>0.22973</td>
<td>0.17558</td>
<td>0.14161</td>
<td>0.11862</td>
<td>0.10924</td>
</tr>
</tbody>
</table>

These numbers demonstrate that while there was a greater percentage of tweets with this convention in 2013, it continues to appear in the corpus tens of thousands of unique times per month and in these five years has only decreased in usage by approximately 1 percent.

5 Mandarin Conventions

5.1 Illocutionary Sentence-Final Modal Particle “～”

There are already sentence final particles in spoken Mandarin\(^3\) that denote speaker attitude (e.g., 吧‘ba’, 呢‘ne’, and 呀‘ya’) that carry over into written

\(^3\) All orthographic representations and transcriptions of Mandarin that occur in this paper use the standard Romanized orthography known as *pinyin* in the standard Beijing dialect created by Zhou Youguang Fox [2017]
language. These can be added as a typed utterance to convey desires, requests for interlocutor confirmation, and other types of attitude (Lu [2005]). However, I argue that in typed Chinese, the sentence-final modal particle “^-” is an additional method of communicating the following: the semantic function of communicating the volition of the speaker, or have the illocutionary effect of soliciting some type of positive speaker interaction. According to informal judgements from a native Mandarin speaker, in response to a suggestion or comment, “_好'hào’” (_yes/good) “_有'yōu’” (_to have) and “_对‘duì’” (_right)” are all considered grammatical usages of [^-] by native speakers that convey this meaning. Example 10 is taken from a Weibo post created on the first of September, 2017.

(10) 我跟女一号@K来个太久
Wǒ gèn nǚ yī hào lái gè tài jiǔ
1.SG with F one number come CLF too long
没有的 自拍 庆祝 一下 ~
méiyǒu de zi pāi qǐngzhù yīxià ~
NEG-have GEN self take celebrate a:little ~

“Me and my number one girl K have gone too long without taking a little celebratory selfie ~”

While a speaker might deduce the intended meaning of this utterance without the addition of “^-” on the end of the sentence, inclusion of “^-” helps to make the speaker’s intentions and emotional state clearer. In-group and out-group distinction is very important in communicating one’s identity and having a sense of belonging to an online community, particularly on Weibo (Han [2015]). In order to create more in-group distinctions, communicating volition and establishing agreement with the interlocutor is important. Without prosody and facial interactions that would normally facilitate this, it appears “^-” has taken on that function.

(11) 有没有好看的雪地靴推荐呀!
Yǒu méiyǒu hǎo kàn de xuědìxuē tuǐjiǎn ya!
have NEG.have good look GEN snowshoes recommend AFFR!
可以带图哦,确定出送一双哒~~
Kěyǐ dài tú ē, quèdìng chūsòng yīshuāng dā ~
Can take picture oh, determine send one.pair AFFR ~~
"Don’t you have good-looking snow boots to recommend! Oh, you can take a picture to determine whether you want to send a pair ~
"

The post above by this user also employs the sentence-final modal particle “~” to elicit agreement, participation, and a sense of in-group belonging (for accompanying picture, see Appendix 11). Here this user is suggesting that the audience look at the various pictures of snowshoes in order to determine whether or not they would want to send a pair to someone. Establishing some sort of connection and ask for agreement will potentially make the interlocutor more likely to buy these shoes and engage with the post, and so use of the sentence-final modal particle not only gives a sense of in-group belonging, but also communicates to any potential interlocutors that this utterance should receive some type of response/interaction.

(12) 卷发 名称 大全，再也 不怕 不知
Juanfā míngchēng dàquán, zài yě bùpà bùzhī
Wavy-hair name inclusive, again also NEG-fear NEG-know
带 怎么 跟 理发 小哥 描述 了 ~
dài zěnme gēn lǐfā xiǎogē miáoshù le ~
bring how with haircut young-man describe PST ~

“With the encyclopedia of wavy hair styles, you’ll never have to be afraid of describing what you bring to the hairdresser!”

Accompanied by a photoset of different pictures of hair (See Appendix 12), this user offers an “encyclopedia” of different pictures of wavy hair for users to take, so that when they go to the hairdresser they can show a picture of the style that they want, rather than not knowing how to describe the style that they like best. Here, the tilde is placed in between the first and second sentences, rather than at the end of the entire utterance. By offering this collection of hairstyles to a public audience, this user is providing a service with the hope that interlocutors use the pictures, as evidenced with the “Do you have one that you like?” comment at the end. In order to more effectively communicate a desire for interlocutor participation, they use the tilde to solicit this and communicate this invitation.
“These red carpet models from last night were too pretty, this color-changing fabric on the skirt lining makes her look almost immortal... No! Oh, she is a goddess no?”

Above, this popular official Weibo account posts about a dress that appeared at an unspecified red carpet event. The user comments that a particular fabric style made the models look ethereal and like an immortal being. Here, the tilde is at the end of the utterance in conjunction with the sentence final particle 哦(ô), which can literally be translated as “oh” or also index that the speaker is suggesting something. This combination of tilde and suggestive particle allows potential interlocutors to know that the user is giving a complimentary suggestion and searching for in-group agreement by indicating that this is not a negative comment.

“In the Weibo post above, this user is discussing older fashion trends that appear to be making a resurgence, suggesting to the audience that trends never really die out. In order to soften this claim and encourage the audience to potentially agree, they added the tilde at the end. Leaving off the tilde would have left the audience with no information about the emotional state...”
for the speaker, leading them to potentially draw the inference that the poster is making a purely factual statement rather than their personal opinion.

While these are just a few examples of this Weibo trend, employing the use of the tilde as a sentence-final modal particle creates multiple meanings when it is added to an utterance. These tildes occur at the end of a whole utterance or sentence, and can co-occur with other sentence final particles that index the emotional state of the speaker. While sometimes the use of the tilde does note the volition of the speaker, it is also used to soften the impact of statements and search for interlocutor agreement and in-group participation. The emotional state of the speaker is often hard to infer without the factors typically present in face-to-face communication, and the tilde communicates to the interlocutors that the utterance does not give rise to the inference that the speaker is upset or commanding, but rather the interlocutor can infer that the speaker is soliciting confirmation from the interlocutor in some way.

5.2 Pinyin Gender-Neutral Pronoun Markers

While there has been some discussion of an emerging gender neutral pronoun into the Chinese orthography (Mair [2013]), not much scholarship has been done on the mechanisms of this shift and its interactions with other linguistic Chinese internet phenomena. When uttered in speech however, the utterance for both third person pronouns is [ta], meaning that in speech the listener cannot always determine the gender of the referent without other context clues. This homophony, or perhaps generality, can be affirming for those who occupy a gender that is outside of the male-female binary or don’t wish for their gender to be mentioned in a third-person context. But this becomes an issue in written Chinese. Written Chinese currently has multiple third person pronouns: one that is male-gendered or indicates the presence of multiple genders(他tä), one that is exclusively female-gendered (她tä), and one that marks inanimate objects (它tä). In the orthography, the distinction is marked with a human radical 人 and female radical 女 respectfully on the left side of the character 他. Historically, cultural interactions with the West in the late 1800’s necessitated the creation of a female pronoun, with 她tä and a different character 伊 fighting for dominance as a female third person singular pronoun until 她 became the most standard use in the 1930’s (黄兴涛 [2009]). In order to account for these identities and/or desire for more gender ambiguity, internet users use the pinyin orthographic system, which
is a Latin-based alphabet that solely transcribes pronunciation Fox [2017].

The example below is an excerpt from a Weibo post made on October 11, 2016. This post advertised a movie about transgender people, and instead of picking a specific written gendered pronoun, the user refers to the subject of the movie without mentioning their gender:

(15) 试图去体会Ta们
zhìtú qù tīhuì Tāmen
attemp go learn-through-experience 3n.pl
所面对 看到了Ta们是
suǒ miànduì kàn dào le Tāmen shì
NMLZ face look-ASP-PST 3n.pl COP
如何积极地面对生活。
rúhé jījìde miànduì shēnghuó.
how positive-ADV face life.

"While they all attempted to learn from their experiences facing hardship and pressure simultaneously, you can also see how they positively dealt with their lives."

Here the user chose to pick the usage that is technically less informative in that it makes the gender of the referents ambiguous, but this ambiguity is not considered to be ungrammatical because it operates under the assumption that the gender of the referents does not need to be explicitly stated, or that their gender is outside of the male/female binary.

(16) 一个 人 的 “原生地” 给ta的
yīgè rén de “yuánshēngdì” gěi tā de
one.CLF person GEN “native area” give 3sg GEN
g气质真是这辈子甩不掉的
qìzhì zhēnshì zhè bèizi dōu shuānbùdiào de
disposition really this life all NEG.lost GEN
士到地心 Low 穿宇宙
tǔ dào dìxīn Low chuān yǔzhòu
soil arrive earth’s-core low pass-through universe

"A person’s ‘native land’ gives their disposition on what their life really is. Both are not lost, going from the soil down low to the
earth’s core, passing through the universe.”

In the above tweet, this user pairs the above status with a picture of students from the show Sailor Moon all raising their hands, while the titular character looks downward, upset. This picture-caption combination indexes onto the emotional state of the speaker in relation to the above utterance. Here, “ta” again replaces 他(tā) and 她(tā) to create more ambiguity around the gender of the person that the speaker is referring to. Since this statement is about the broad category of people instead of a specific person, creating more ambiguity around the gender of the speaker is more informative in that the interlocutor can apply this statement and subsequent feeling to many people. This also serves to make the status more relatable, and more likely to be liked or reposted.

While one of the Weibo posts is explicitly about people with genders outside of current societal norms, the other is a more mainstream usage that does not have anything to do explicitly with these genders, but is simply using this convention because of the ambiguity that it creates. While it is hard to draw more definitive conclusions with these data given the lack of a monolithic Weibo corpus, generic searches of ta using Weibo’s search function produces over fifty pages of recent results containing this term. The written form of this expression is curiously more informative than the spoken form. Written communication does not always have a standardized method for conveying prosody and other informative features, making speaker identity harder to parse. In spoken Mandarin, however, the spoken form lacks a normalized, specified orthographic convention (other than perhaps gesture or context) to disambiguate gender. The emergence of this convention appears to be in response to this disambiguation and imposed binary. In order to account for non-binary identities and keep the gendered ambiguity of spoken Mandarin available in written Mandarin, internet users simply use the pinyin of the character, since this keyboard is already available to them and pronunciation would remain consistent.

6 Discussion

Other studies analyzing phenomena such as digitally written intensifiers and initialisms find that users, particularly young adults, navigate internet communication at many different registers with ease and maintain stable patterns
with these usages (Tagliamonte [2016]). This is consistent with the new phe-
omena observed in this paper—speakers can reproduce and deduce meaning
from emerging conventions. When the meanings are not always apparent to
a user, they can go on message boards where the community informally dis-
cusses grammatical usages of these conventions and their meanings. For
those who want to acquire new internet vocabulary, these act as living dic-
tionaries. Once meaning is acquired, users then put them into their internet
vocabulary and use them in their own utterances to circumvent problems in
deducing the semantic meaning.

As is the case with emerging conventions that have not fully been inte-
grated into the lexicon, their meanings must be defined and disseminated.
Many of these conventions began within smaller groups of the internet (such
as people who enjoy the music group Vaporwave and simulate the album text
style), and spread in reach to other internet communities. Notably, just like
spoken conventions, these written conventions are not prescribed by govern-
ing bodies, which allow for communities to define their meaning. For internet
users eager to understand these emerging conventions and incorporate them
into their own lexicon, discussion threads and internet dictionaries, most no-
tably the forum site Reddit, Urban Dictionary, and Know Your Meme are an
invaluable source. These conversation threads investigate the origins of con-
ventions, popular usages, and meanings. For example, there is an archived
thread on Reddit about the use of “TM”, calling the convention a “sarcastic
trademark” (red [2015b]). At the time of the thread (2015), “TM” was being
used by a small group of users who paired it with the host “Nice Guys” to
describe and make fun of men who view themselves as “caring and emphatic
gentlemen in courtship to increase their potential towards women... mainly
in anticipation of a relationship or sexual favors in return”, and then use of
this convention expanded to include most internet-savvy users (“What does
the overhead TM mean” [2015]). Suprasegmental spacing, according to both
Reddit and Know Your Meme threads first emerged in fan communities of
the group Vaporwave looking to mimic the aesthetic of their album covers,
which feature words with spaced out letters. These forums posit that the
convention first came from that community, spread to people “making light-
hearted fun of how serious some people take a musical genre that is in no way
supposed to be taken seriously”, and from that point was known by most
of the internet that it began to be used among most internet-savvy users
(“Know Your Meme” [2017] and “What is up with a e s t h e t i c?” [2015]).
These is of course the possibility that this convention could have taken on the
same meaning some other way, but unofficial recordings and discussions of
the convention as well as my own analysis of the most current meaning lead
me to posit that its etymology had something to do with the formation of
the current meaning. "~/" has had a bit more mainstream examination of
its meaning and user base, with a New York Magazine article (Leslie [2017])
citing the Twitter community as the ones most likely to use this convention,
which is confirmed by my own corpus as the most used convention out of the
three observed on Twitter. More recently, however, the communities that
use this convention now include Facebook users and even popular magazines
(e.g., Cosmo) aimed at demographics that use the internet frequently, as
shown in this article by Bernstein [2015]. The tilde in Mandarin Chinese
appears to denote a similarly broad internet community membership, with
many users employing this convention on everything from emails to texts
to social media posts4 ("Meaning of ‘~/’ at the end of a sentence" [2013]).
The use of ta in Chinese, however, signifies a more narrow community mem-
bership than the aforementioned conventions. As previously discussed, use
of ta seems to note that the mostly young users of this convention have a
particular stance towards gender in a few possible ways: both existing third
person pronouns in Chinese do not account for non-binary genders, and/or
historical use of 他(tā) to refer to multiple genders does not currently convey
the desired gender ambiguity. While this phenomenon has been observed in
cases as far back as 2013, use of this convention still has not appeared to
have been integrated into larger Chinese internet communities (Mair [2013]).

The natural meanings of these conventions arise from metalinguistic know-
ledge about who uses these conventions most on the internet. Users who have
the ability to felicitously use these conventions demonstrate internet user in-
group knowledge and (unofficial) membership. Regarding both English and
Chinese conventions, the natural meaning (Grice [1957]) conveyed through
their usage is that the user is fairly internet-savvy, and most likely ranges in
age from teenager to young adult, as these demographics are the most likely
to use the internet now and have grown up using and adapting this medium
to communicate (Anderson and Perrin [2016] and "Breakdown of Microblog
Users in China" [2017]). In addition, use of these conventions also conveys
that the identity of the user is most likely a regular internet user in that

4There is even discussion of this “non-standard” usage even being incorporated into
mainstream punctuation by appearing in the more prestigious Chinese National Standard
for the Usage of Punctuation (Mair [2013]).
age group, as opposed to someone who is running an account for a business, corporation, or group. Successful use of these conventions also demonstrates that the user understands and/or is a part of various internet and youth cultures. Because these conventions carry these natural meanings, occasionally companies will attempt to use them in their branding to market to young internet users. But as Grossman [2014] notes, they are often infelicitous in their usage and thus unsuccessful, prompting Twitter users to make fun of business accounts attempting to interact with a group they are not a part of for capitalistic gain.

The internet allows for these novel methods of identity exploration, partially because of this globalized community and ease of communicating with individuals despite physical distance. Thus, when problems such as communicating without prosody arise, speakers find novel ways of reproducing these signs on the internet, and other users quickly catch on and mimic the convention until it gains viral popularity and enters the internet lexicon on a large scale. Much of this communication occurs through discourse on various online platforms whose whole purpose is widespread dialogue, such as microblogs. Users view choices in utterance and convention as “accurately reflecting [their] ‘real-life’ identity, and of regional or social variation in language use” (Wikström [2016]). These blogs, which are repositories for these utterances, represent a user’s online identity and thus the language they use constructs a “consistent public face” to represent themselves along the lines of gender, sexual orientation, race, etc. (Huffaker and Calvert [2005]). In order to accurately represent themselves and find community online, speakers use trends present in specific speech communities and apply them to their own lexicon, such as the Scottish tweet in the example given by user B in example 4, which uses a convention present in other modes of English internet communication.

7 Conclusion

These five conventions demonstrate that the speaker’s emotional state, prosody, and other identity-based factors can be communicated in multiple languages despite the lack of a previous standardized written conventions for these phenomena. Being able to convey all communicative aspects in this medium allows users to more accurately represent themselves online, constructing an online persona that is representative of their own real life identity. These
conventions reduce ambiguities that have long been present in written communication, but this new globalized space has helped to facilitate this language change. The conventions that arise from these ambiguities are not segregated to one singular language or dialect, but as the data show are present in multiple languages and dialects across the world.

Constructing an accurate online representation of the self can help users find and maintain community. Creating these new conventions also helps to reduce the communicative burden for both sides and ensure more effective discourse. Blogs and the internet as a whole represent changing modalities of communication, and these modes act as mirrors for speakers to view not only themselves but the billions of others in this globalized space as well. In order to adjust to the growing popularity of internet communication, new conventions arise, and it is imperative that these conventions are documented in order to observe language change and the adaptability of speakers.
8 Appendix

Example 3

Example 4

My heid is an erratic mess of ideas. Constant. And no just with food. Just constant fucking mental overload. It's why I like Twitter.

that's livin ™

2:43 PM - 1 Jun 2015

1 Like
Example 5

real shit tho Tyler the Creator is a rly important creator to me n so is Kanye and ik they're both ~problematic~ so. don't bother telling me

2:53 PM - 30 Jun 2017

Example 6

i accidentally saw your sekai tweetfic and sobs loudly because too ~cuuute~ my mouth hurts from smiling too much ;A;

1:48 AM - 25 Jun 2014

Example 7

are chris rock and "the rock" related?

um, is that a real question

im offended you thought i was serious

It's called.. wait for it...

~"sarcasm"~

8:24 PM - 27 Jun 2013
Example 8

Ok y'all lets get something straight Black Panther is a BLACK movie not a "POC" movie. It is a BLACK movie.

Example 9

My shift is dragging

Example 11

有没有好看的雪地靴推荐呀！可以带图哦，确定出送一双哒～～
Example 12

Example 13
Example 14

11·2 10:00 come from weibo.com
110年前的欧洲女孩。时尚是个轮回~

Example 16

10·16日 00:55 来自 iPhone 7 Plus
一个人的“原生地”给ta的气质真是这辈子都甩不掉的，土到地心，Low穿宇宙
References


What is up with aesthetic?, 2015a. URL https://www.reddit.com/r/OutOfTheLoop/comments/3snn9u/what_is_up_with_a_e_s_t_h_e_t_i_c/.

What does overhead tm mean, 2015b. URL https://www.reddit.com/r/OutOfTheLoop/comments/3qjdqj/what_does_the_overhead_tm_mean/.


Weibo celebratory selfie post, September 2017e. URL https://weibo.com/1660399881/Fj02Id65v?type=comment#_rnd1511028760416.


Shaoyong Chen, Huanming Zhang, Min Lin, and Shuanghuan Lv. Comparison of microblogging service between sina weibo and twitter. 2011.


32


Jess Kimball Leslie. The internet tilde perfectly conveys something we don’t have the words to explain, June 2017. URL http://nymag.com/scienceofus/article/why-the-internet-tilde-is-our-most-perfect-tool-for-snark.html.


