The Effects of Grammaticalized Possibility and Evidentiality on the Epistemic Responsibility of Interlocutors, in Tweets

Matt Rabitsky, Tahvo Wilson

University of British Columbia

Abstract

In a world of increasing misinformation and distrust amongst speakers, it can prove insightful to analyze the language the structures that are used by those who are being intentionally responsible (or irresponsible) to the truth in their speech. To this end, this study used a proprietary codebase to collect and analyze tweets that exhibit certain grammatical structures that are ostensibly pertinent to misinformation and responsibility to knowledge. While no strong correlative links were found, a *post mortem* of the study revealed clear errors in the methodology that, if corrected, could set the groundwork for a potential explosion of research in the area of cognitive implications of modal language in popular vernaculars, cross-linguistically.

Keywords: Evidentiality, possibility, tweets, natural language processing

Introduction

While there already exists much grounding work on the existence of evidentials and possibility markers in many languages (e.g. Tenny, 2006; Wiemer, 2007), as well as studies related to the acquisition and use of such morphological structures by children of varying ages (e.g. Fitneva, 2009; Ozturk & Papafragou, 2008), there is a gap in the academic discussion regarding the impact of these grammatical forms on everyday conversation, in particular on the digital platform of Twitter. Our research addresses this lack in knowledge by examining only tweets which contain markers of evidentiality and possibility. We evaluate these tweets for their sentiment value and weight in the twitter community, through likes and retweets: research has recently been conducted on the emotion density in tweets in English and Arabic (e.g. Pribán, Hercig & Lenc, 2018); however, we wish to apply these same principles to our set of languages, including German, French, Japanese, and Korean. We begin by discussing the importance of evidentiality and the impact it has on the epistemic responsibility — more simply put: the credibility — of the speaker. This is followed by a detailed discussion of our data collection in which we investigate the impact that evidentiality and possibility encoding have on the epistemic responsibility (*q.v.* Appendix A) of everyday speakers, using specific examples.

Background. Evidentiality — a grammatical structure pertaining to the expression of direct/indirect evidence known by the speaker (de Haan, 2013) — and possibility — an expression of either the situational or epistemic probability of the given information (van der Auwera, Johan & Ammann, Andreas; 2013) — are two interesting morphological features that exist in one form or another in many languages. They carry the potential to inform and encode certain information in a dialogue that is not clearly stated in the content, but rather speaks to the

trustworthiness or veracity of the information being discussed. We believe that because of the unique nature of these markers, and their relative ubiquity in languages across the globe, these two features have the ability to impact the epistemic responsibility of those who use them in twitter. Indeed, there already exists some evidence that children who understand these markers are able to use them effectively to make choices regarding trust in sentential information (Fitneva, 2009). This use of evidentiality and possibility leads to a high sentiment value and relatively high number retweets and likes which is discussed later in the paper.

Methodology & Data Collection

With a character limit of 240 characters and large number of users, Twitter is *the* model for today's fast-paced, provocative news and communication. For this reason we thought that analyzing tweets would provide an accurate representation of the use of evidentiality and possibility in everyday speech in the aforementioned languages. Our corpus consists of one thousand tweets from each of the four languages, French, German, Japanese, and Korean, coming from randomly selected Twitter accounts, to provide an unbiased representation of the use of epistemic markers throughout the language.

Methodology. The tweets, and analyses thereof, were completed programmatically (*q.v.* Appendix B), following a procedure that can be generally described as such: download one hundred tweets in each language[1], which must contain the appropriate grammatical markers for that language (*q.v.* Appendix C); perform several natural language processing (NLP) analyses on the data (*viz.* document similarity, sentiment analysis, lemmatization); take the numerical values from the previous NLP analyses and combine them with other numerical values produced

[1] The Japanese and Korean tweets have the additional limitation of being only from the last week (this is due to a constraint in Twitter's API, and is not an intention of the authors).

from the tweets (*viz.* the number of likes, the number of retweets, the number of elements of each syntactic category) into a tabular format, upon which we then perform unsupervised learning to suss out the clusters and patterns present in the data; finally, having found patterns in the data, as well as having the results of the previous NLP analyses, we plot and visualize the data appropriately, so as to be able to draw relevant conclusions.

Assessments and Measures. Epistemic responsibility is inarguably an abstract and difficult to measure quantity; however, when prompted, most people will be able to qualify whether an interlocutor is being epistemically responsible, or if they are being willfully ignorant. As it stands, insufficient research has been conducted (to the researchers' knowledge) on what these qualia specifically are, and how they might be measured in a meaningful way; thus, rather than speaking to whether any particular feature of the dataset has a direct consequence on the epistemic responsibility of any one interlocutor, we will be treating any significant correlation as a meaningful affect.

Results

Due to the number of features present in the dataset (*q.v.* Appendix D), it was necessary to reduce the number of dimensions in which the data existed down to two, as otherwise they could not be presented graphically. This was achieved through the use of three different dimensionality reduction algorithms, namely and in order of importance: t-SNE, MDS, and PCA. After reducing the dimensions of the data, three different clustering algorithms were applied, to see if there were any statistically significant clusters in the dataset, wherefrom we could dig deeper and actually read the tweets to try to understand the nature of the similarities. Below are the graphs showing the results of these operations on the data.

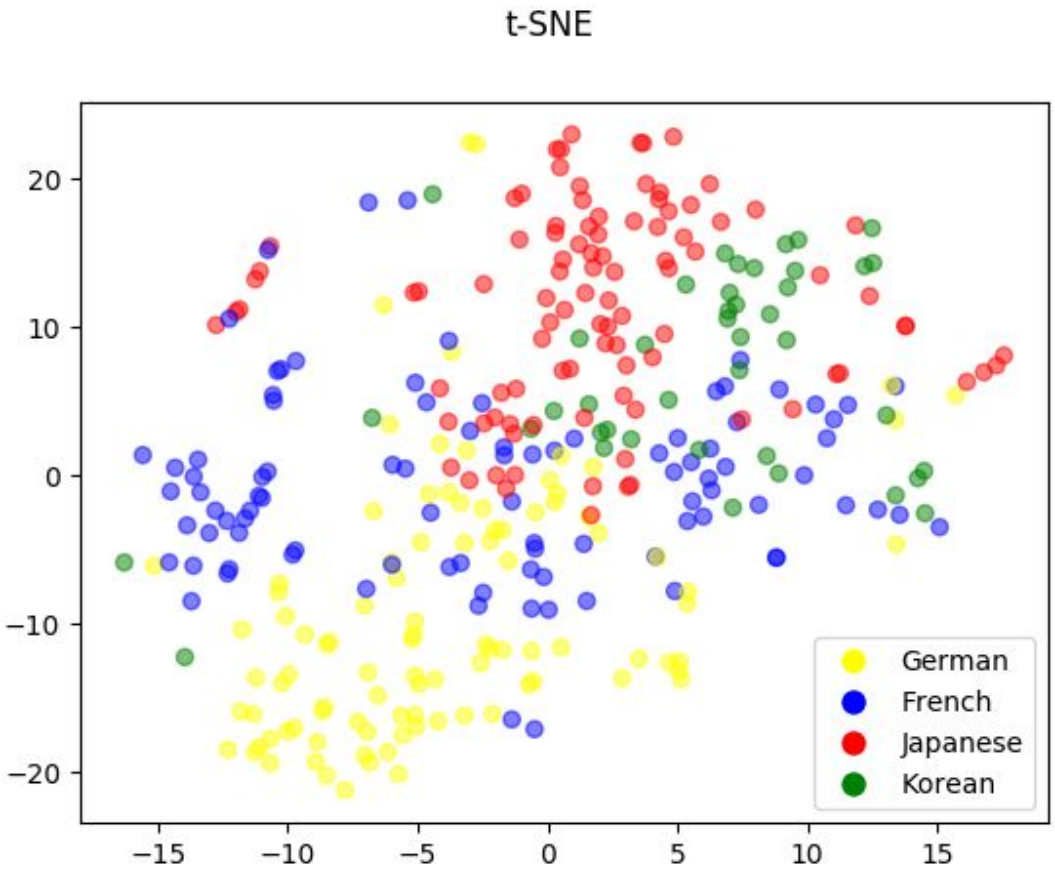

Figure 1.1 t-SNE.

As a note, because the t-SNE algorithm has a random initialization, this graph looks different than the one immediately below it; however, the general gist remains the same. In both cases, t-SNE did not provide particularly promising results, as there were not really any significant clusters to be found.

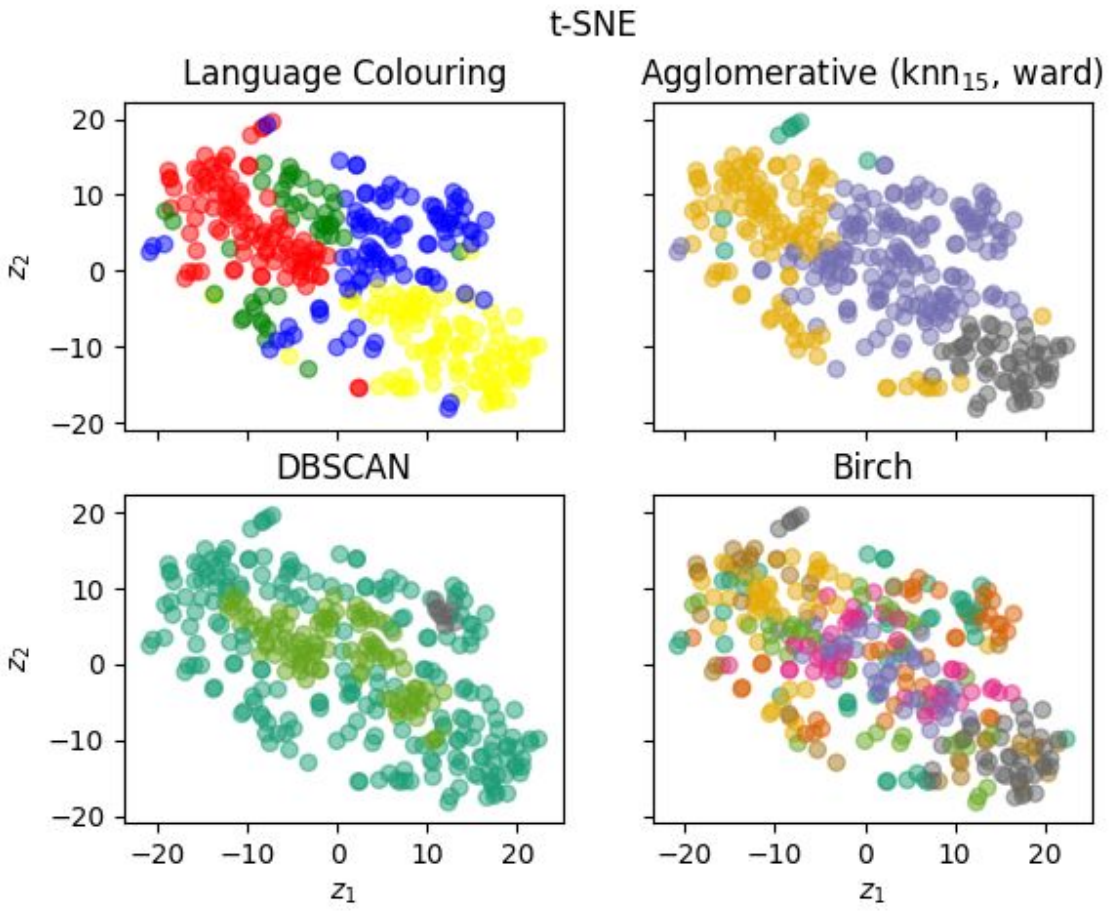

Figure 1.2 t-SNE and associated clusterings.

t-SNE did not provide any particularly meaningful insights into the dataset, unfortunately. It would appear that the entire dataset was distributed rather evenly in a diagonal blob of sorts. The density-based clustering (DBSCAN) found the inside to be more clustered than the outside; however, this was not reflected strongly in the other clustering algorithms.

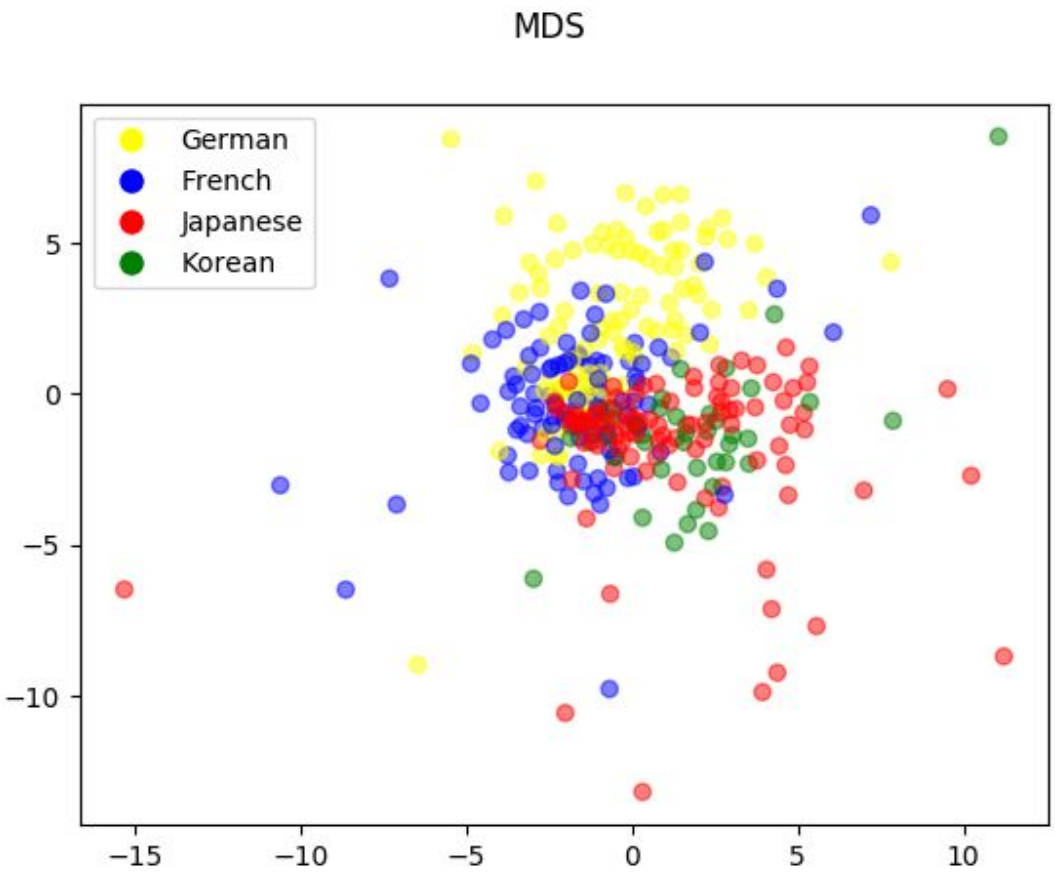

Figure 2.1 MDS.

Unfortunately, MDS proved even less useful than the t-SNE interpretations, as here the data is simply clumped all together in a central ball of sorts, have a smaller spread than the t-SNE exhibited by about a factor of half.

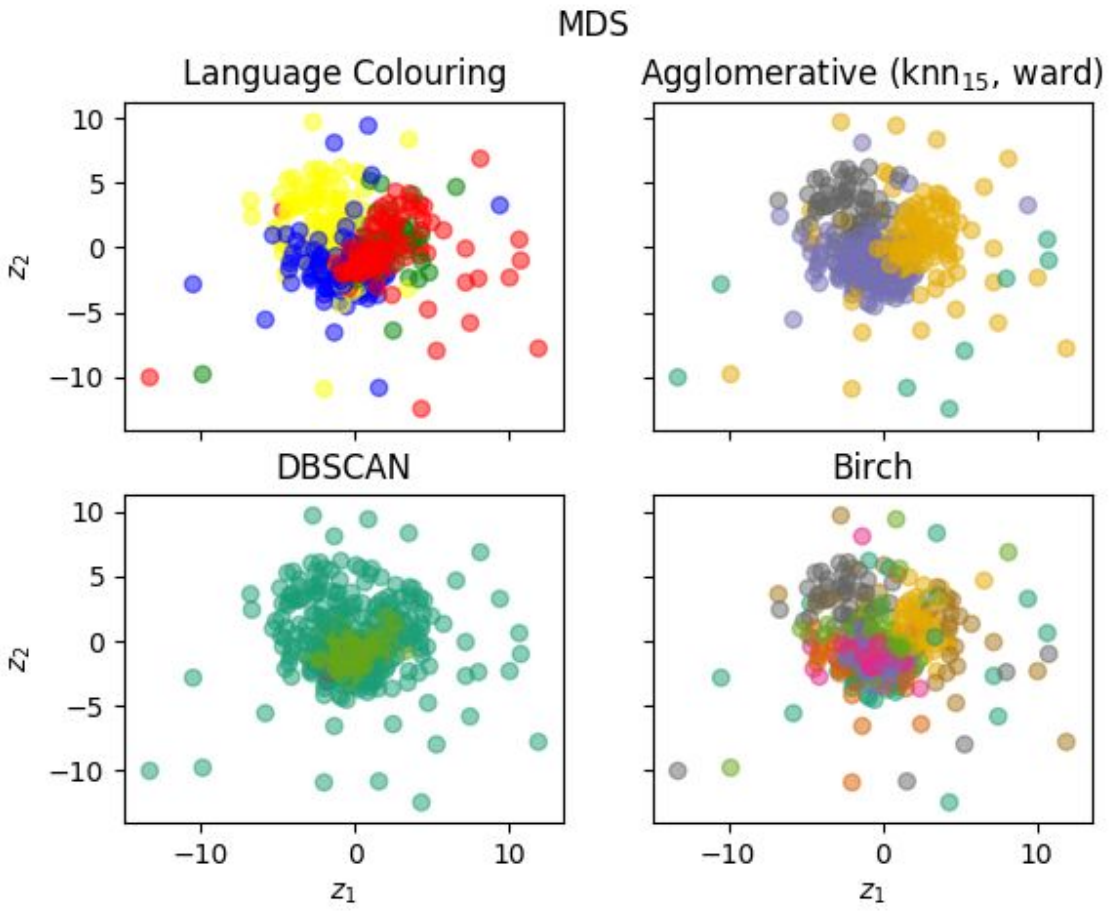

Figure 2.2 MDS with associated clusterings.

The clusterings for MDS did not help, because although the density cluster showed a higher density in the centre of the ball, the agglomerative clustering (which takes into account nearest neighbours for every datapoint) clustered along language lines, dissuading us from perceiving a significant importance (this was fairly well reflected in the Birch clusterings as well, falling more or less along the language lines).

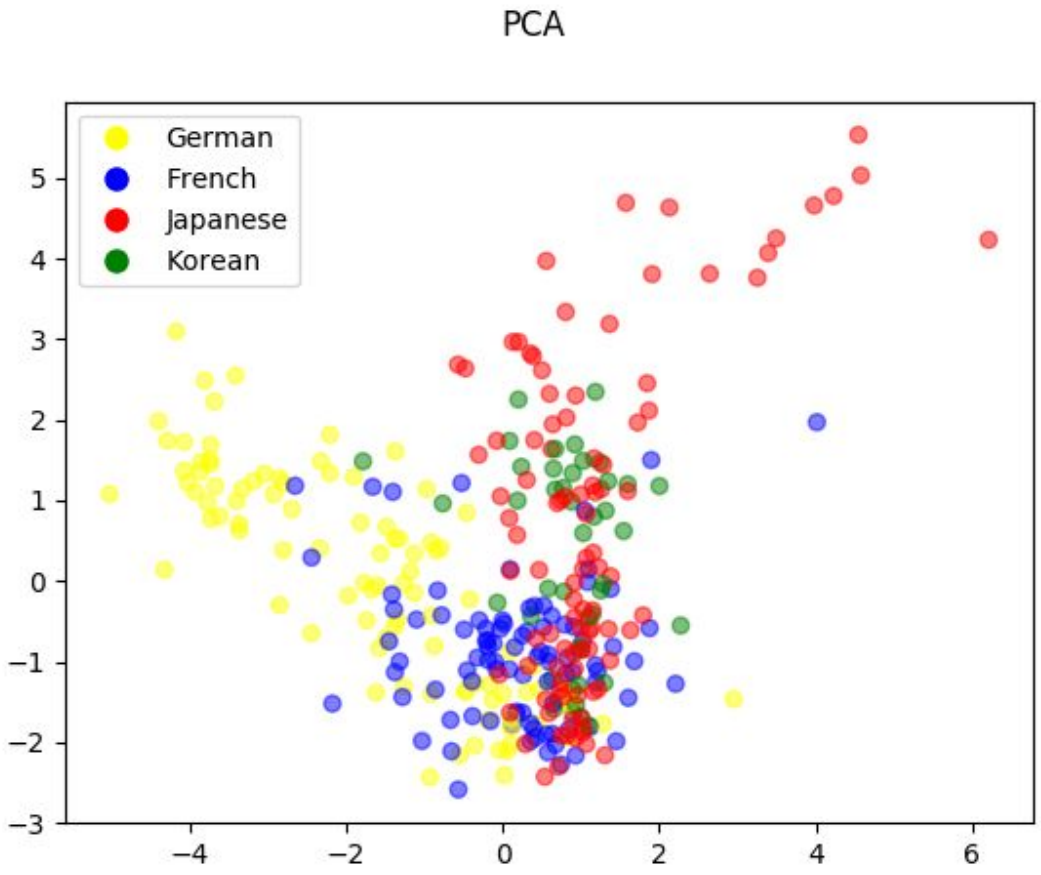

Figure 3.1 PCA.

Of the three dimensionality reduction algorithms, PCA seemed most promising, as it seemed to clearly depict a convergence of the dataset into one cluster (roughly at the point (0, -1.5)) with several tails.

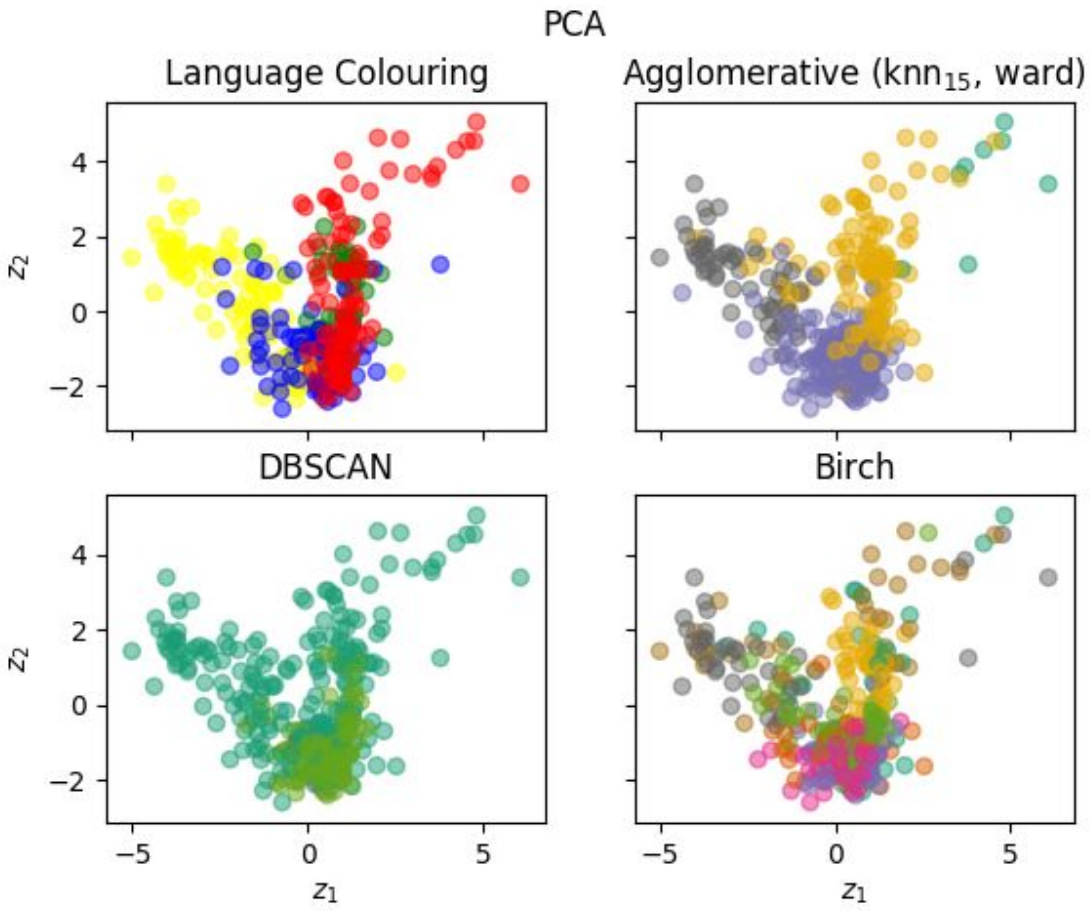

Figure 3.2 PCA with associated clusterings.

The clusterings showed that our initial visual assumptions were true regarding the density of the bottom section. However, the Birch and agglomerative clusterings showed that despite the seeming progress that PCA had provided, the datapoints were still strongly tied to their respective languages in an uneven distribution. In this, we found our apparent pitfall.

Discussion

In an attempt to combat this perceived linguistic dependence, we tried a second batch of PCA on the same dataset, but with the columns containing the language information removed. This did not prove to be particularly distinct from the previous PCA attempt, showing that the

actual, explicit language information was not a strongly contributing factor to the principle components:

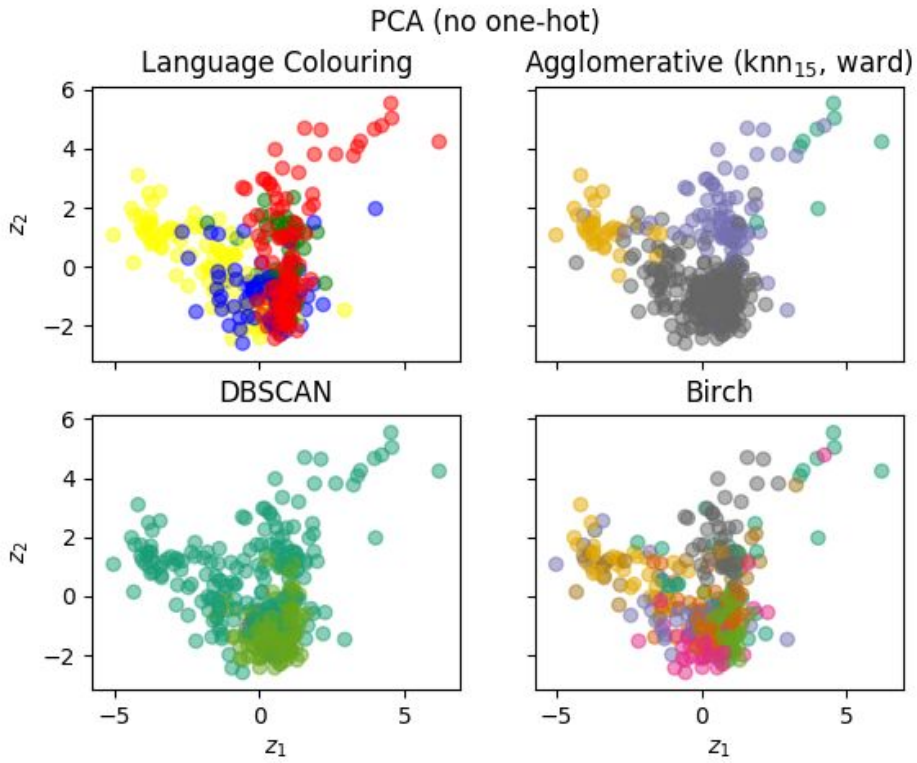

Figure 4 PCA without the one-hot encoded language data (i.e. without the first four columns)

The agglomerative clustering in this particular case seemed to be more indicative of a cross-linguistic hotspot at the bottom of the figure, but this was not well reflected in the Birch clustering, or in the initial visual analysis (the German and Japanese 'tails' still existed).

We know from individual analyses of some of the tweets that use of these grammatical markers implied an impact upon the epistemic responsibility of the author, as shown in the case below; but, we feel that this could not accurately be represented in the data, as it has been presented.

1113546471985561605

Beim Standard scheinen[2] vernünftige und denkende Menschen zu arbeiten[2] (also auf jeden Fall vernünftiger, als bei der Springer-Presse). Der Autor rät mit Blick auf die Ukraine-Wahl, die "Ost-West-Brille" abzunehmen.

By default, reasonable and thinking people seem to work (in any case, [they're] more sensible than the Springer press). The author advises with a view to the Ukraine election to take off the "East-West glasses".

In this particular tweet, the author seems to be lampooning the linked Springer Press article regarding the Ukrainian civil war and election. In context[3], however, it can be found that this particular author is using the indirect evidential in an effortful attempt to mislead the reader: this particular tweet is authored by a journalist associated with Russian state propaganda, and thus has a particular interest in misrepresenting the Ukrainian situation; so, it is apparent that their use implies an intentional epistemic dishonesty.

In all three attempts at dimensionality reduction, despite the density-based clustering having shown some similarity between the four languages (this language similarity being the goal of the study, as a cross-linguistic clustering would imply some additional value outside the scope of any one language), the other two clustering algorithms displayed a greater affinity for the language-specific groupings. This, in combination with the individual tweet readings, would lead us to believe that while there might be something worth looking at here regarding cross-linguistic implications for epistemic responsibility, the features we chose to operate on

[2] Researchers' emphasis, not author's.

[3] We did not feel it would be appropriate to name the tweet or the author; however, all of the tweets in the dataset are publically available, and can be looked up based on the tweet IDs provided, should the reader choose to do so.

were not sufficient (or possibly even necessary) to definitively say anything one way or the other: in short, the language-specific clusterings prevent us from making broad claims because we cannot show independence between the variables.

Conclusions

Epistemic responsibility is a rather nebulous concept, and it would seem that we have found that it is not a concept easily made numerical. Having encountered particular examples where there was an effect on the epistemic responsibility of the interlocutor, and having seen glimmers of promise in the visualized data, we think we may be on to something. However, it is equally apparent to us that our methodology was severely flawed, and that much further research needs to be conducted in the area before and further lookings can be had at the idea, or any bountiful conclusions drawn.

The method of using syntactic categories proved to be insufficient with regards to drawing distinctions between the language groupings. While we had originally intended to use more salient features, such as whether or not the tweet contained a question, we found that there were too many problems that could not be solved in a reasonable time as to allow this paper to be completed (as the example, determining whether a tweet contains a question or not is particularly difficult, in that punctuation in everyday speech is not always used as we might hope it is, and would likely require the training of a machine learning model to classify tweets based on their word choice or somesuch). Further research would need to be conducted, likely in the area of experimental philosophy, to elicit more promising features: as we mentioned previously, most rational humans can effectively attribute whether an agent is being epistemically responsible or not; however, the quantification of that gut reaction would be required in order to operate on

such data *en masse*. It is also important to note that we did not perform a comparative study between tweets that did and did not contain these modal markers, as doing such a comparison was outside the scope and means of this study. That being said, a comparative study would potentially draw more attention to the impact of those modal markers: as it stands, we are only looking at correlative measures, and with those alone we could not show a causative force, as we might with a comparative study.

In all of this, while we did not find exactly what we were looking for, we feel that we have set an appropriate entry point into the discussion on the implications of modality on human cognition and intention.

References

van der Auwera, Johan & Ammann, Andreas. (2013). Situational Possibility. Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/74>

van der Auwera, Johan & Ammann, Andreas. (2013). Epistemic Possibility. Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/75>

Fitneva, S. A. (2009). Evidentiality and trust: The effect of informational goals. *New directions for child and adolescent development*, *2009*(125), 49-62.

de Haan, Ferdinand. (2013). Semantic Distinctions of Evidentiality. Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://wals.info/chapter/77>

McHugh, C. (2013). Epistemic responsibility and doxastic agency. *Philosophical Issues*, *23*, 132-157.

Ozturk, O. & Papafragou, A. (2008). The acquisition of evidentiality in Turkish. *University of Pennsylvania Working Papers in Linguistics*, *14*(1), 23.

Papafragou, A., Li, P., Choi, Y., & Han, C. H. (2007). Evidentiality in language and cognition. *Cognition*, *103*(2), 253-299.

Pribán, P., Hercig, T., & Lenc, L. (2018) Emotion Intensity Detection in Tweets. *NTIS – New Technologies for the Information Society.*

Wiemer, B. (2007). Lexical markers of evidentiality in Lithuanian. *Rivista di Linguistica*, *19*(1), 173-208.

References: Corpus

Dendale, Patrick & Van Bogaert, Julie. (2007). A semantic description of French lexical evidential markers and the classification of evidentials. Italian Journal of Linguistics. 19.

Diewald, G., & Smirnova, E. (2010). *Evidentiality in German: Linguistic realization and regularities in grammaticalization* (Vol. 228). Walter de Gruyter.

Lee, C. (2010). Evidentials and Epistemic Modal in Korean: Evidence from their Intractions. In *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*.

McCready, E., & Ogata, N. (2007). Evidentiality, modality and probability. *Linguistics and Philosophy*, *30*(2), 147-206.

Miche, E., & Lorda, C. U. (2014). Probability and certainty markers in French and in Spanish (sans doute/sin duda). *Language and Dialogue*, *4*(1), 42-57.

Patrick, D., & Van Bogaert, J. (2007). A semantic description of French lexical evidential markers and the classification of evidentials. *Rivista di Linguistica*, *19*(1), 65-89.

Song, K. A. (2010). Various evidentials in korean. In *Proceedings of the 24th pacific asia conference on language, information and computation*.

Tenny, C. L. (2006). Evidentiality, experiencers, and the syntax of sentience in Japanese. *Journal of East Asian Linguistics*, *15*(3), 245.

Wymann, A. T. (1996). *The expression of modality in Korean* (Doctoral dissertation, Universitat Bern).

Zifonun, G., Hoffmann, L., Strecker, B., & Ballweg, J. (1997). *Grammatik der deutschen Sprache* (Vol. 1). Walter de Gruyter.

References: Code Packages

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.".

Park, E. L., & Cho, S. (2014, October). KoNLPy: Korean natural language processing in Python. In *Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology* (Vol. 6).

Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In *Proceedings of HLT-NAACL 2003*, pp. 252-259.

Appendix A

Epistemic Responsibility

Epistemic responsibility is the notion that we as humans are responsible for what beliefs we have, and that we may be considered blameworthy for expressing unjustified or untrue beliefs (whether or not they are posed fraudulently).

Epistemic responsibility is not much different from the standard sort of responsibility: if you are assigned a task and you complete it in an unreasonable manner, as is the case if you are asked to wash dishes and you shatter them all with a pressure washer, then you did not act responsibly; similarly, if you are assigned a task but are not able to complete it due to forces outside of your control, as in the case where, having been tasked again with washing the dishes, your sibling comes in and smashes all the dishes before you can clean them, then you cannot be held responsible (*i.e.* liable) for the task being incomplete. It may be considered that anyone who wishes to engage in reasonable discourse is tasked with doing so in a manner that is appropriate and effective, and so we can and must hold ourselves epistemically responsible in such cases.

As a note, it must be understood that the notion of epistemic responsibility hinges on the presupposition of doxastic voluntarism — that is, the premise that we choose our own beliefs (one way or another). If we accept our own doxastic agency, and expect the same from others, then we may assign epistemic responsibility (McHugh, 2013).

Appendix B

Code

The full code can be found at <https://github.com/mRabitsky/WRDS150-corpus-constructor>. Within the code itself, there are numerous comments explaining the step-by-step process that was followed in order to collect and operate upon the data. A more generous explanation thereof can be found in the README.

Appendix C

Possibility & Evidentiality Markers

Each of the words/morphemes in the second table below imply either the possibility or necessity of their parent preposition, or they show that the speaker has certain evidence to support the preposition being stated. In order to make this more clear, they are marked with one of the six codes from the first table below to make it easier for you, the reader, to understand the usage thereof.

	Deontological	Epistemic
Possibility	**DP**	**EP**
Necessity	**DN**	**EN**

	Evidential
Direct	**VD**
Indirect	**VI**

Table C.1 Modal codes

Here, below, is a table listing all of the search terms used within the project, per language. The table explains more or less exactly the construction that is expected (as certain elements have multiple meanings and only carry their modal meaning in specific contexts); however, for an exact understanding, it would be best to see the code to find the specifics.

Language	**Markers**		
French	(Conjugations of) the verbs [*croire*], [*pouvoir*], [*trouver*]; the phrase [*sans doute*].		
	croire	**EP & VI**	‘to believe’
	pouvoir	**DP**	‘to be able to do [something]’
	trouver	**EN & VD**	‘to find’
	sans doute	**DP**	‘without a doubt’

<table>
<tr><td></td><td>(cf. Dendale & Van Bogaert, 2007; Miche & Lorda, 2014)</td></tr>
<tr><td>German</td><td>(Conjugations of) [drohen], [versprechen], [scheinen] along with a [zu]-infinitive;
(Conjugations of) [werden] along with an infinitive verb.
<table>
<tr><td>drohen</td><td>VI</td><td>‘threaten’</td></tr>
<tr><td>scheinen</td><td>VI</td><td>‘seem’ / ‘seem to be’</td></tr>
<tr><td>versprechen</td><td>VI</td><td>‘promise’</td></tr>
<tr><td>werden</td><td>VI</td><td>‘become’</td></tr>
<tr><td>zu</td><td>~</td><td>‘to’ (i.e. [zu haben], ‘to have’)</td></tr>
</table>
(cf. Zifonun, Hoffmann, Strecker, & Ballweg, 1997; Diewald & Smirnova 2010).</td></tr>
<tr><td>Japanese</td><td>[みたい] (‘mitai’), [らしい] (‘rashī’), [そう(だ)] (‘sō(da)’), [よう(だ)] (‘yō(da)’).

All four of these morphemes mean roughly the same thing, though they are quite hard to translate into English as their usage is deeply rooted in the milieu of Japanese grammar.

Each morpheme can be understood to mean “like” or “seem”, but in reality each act as a comparative particle. [みたい] and [よう] are used to compare the agent of the sentence to some other quality, trait, or idea. [らしい], on the other hand, affords an indirect inferential modality to the entire sentence in which it appears (cf. McCready & Ogata, 2007).</td></tr>
<tr><td>Korean</td><td>[책무 + 이다] (‘chaengmu + i-da (copula)’),
[허가] + ([있(다)] or [없(다)] or [얻(다)] or [받(다)] or [되(다)])
└ (‘heoga + (iss(-da) or eobs(-da) or eod(-da) or bad(-da) or doe(-da))’),
[허락] + VERB/AUX (‘heolag’), [허용] + VERB/AUX (‘heoyong’),
[—면 + 좋(다)] (‘—myeon & joh(da)’),
[필요 + ([있(다)] or [없(다)] or [하다])]
└ (‘p^hil-yo & (iss(-da) or eobs(-da) or ha(-da))’),
[—도 좋(다)] (‘—do joh(-da)’), [—야 하다] (‘—ya ha-da’),
[추측] (‘chucheug’), [추정] (‘chujeong’),
[—걸] (‘—geol’), [것 같다] (‘geos gat^h-da’),
[—것다] (‘—geos-da’), [판단] (‘p^handan’), [생각] (‘saeng-gag’),
[상상] (‘sangsang’), [의견으로] (‘uigyeon-eulo’),</td></tr>
</table>

[—군] (‘—gun’), [—네] (‘—ne’), [—다고 하(다)] (‘—dago ha(-da)’).

책무	**DN**	‘duty’, ‘obligation’, ‘accountability’
이다	~	‘to be’, the Korean existential copula
허가	**DP**	‘permission’
있다	~	‘exist’
없다	~	‘exist.NEG’ (i.e. the negative variant of the previous morpheme)
얻다	~	‘receive’, ‘get’, ‘obtain’
받다	~	‘receive’, ‘undergo’, ‘catch’
되다	~	‘become’
허락	**DP**	‘allow’, ‘permission’
허용	**DP**	‘permit’
—면	**DP**	‘if’, suffix indicating the conditional conjunctive mood
좋다	**DP**	‘to be good’
필요	**DN**	‘necessity’, ‘requirement’, ‘need’
하다	~	‘do’
—도	**DP**	Postposition additive particle, implying emphasis
—야	**DN**	Deontic modal suffix (when used with [하다])
추측	**EP**	‘surmise’, ‘guess’
추정	**EP**	‘presumption’, ‘deduction’
—걸	**EN**	Suffix denoting epistemic necessity, grammaticalized from [—것 + 이다]
–것–	**EN**	Traditional future-tense infix marker, but can be interpreted as a modal
것	**EN & EP**	‘thing’

같다	**EP**	‘seem’
—것다	**EN**	*see similar suffix above, —걸; both are grammaticalized encodings from the same construction*
판단	**EN**	‘judgement’
생각	**EP**	‘thought’, ‘thinking’, ‘idea’, ‘concept’
상상	**EP**	‘supposition’, ‘imagination’, ‘assumption’
의견	**EN**	‘point of view’, ‘opinion’; (always affixed with topic marker [으로])
—군	**VD**	Direct evidential suffix in the informal and neutral moods
—네	**VD**	Direct evidential suffix in the informal and neutral moods, with a greater level of perceived veracity than [—군]
—다고	**VI**	Hearsay evidential

(*cf*. Wymann, 1996).

Table C.2 Grammatical Markers Per Language

Appendix D

Data

Subsequent to this page, please find a tabular rendition of part of the data. Due to the size of the data, it would be impertinent and inconvenient to include it all; however, all of the data is available, in its totality, in the CSV document online, in the GitHub repository.

A Brief Description of the Data. The tweets that we collected we processed into tabular rows in a 25 column table, with each column representing a particular 'feature'[4] of the data. These 25 columns were as follows:

Feature Name	**Description**
id	The ID of the tweet (the unique number assigned by Twitter)
de	1 if the tweet is in German, 0 otherwise
fr	1 if the tweet is in French, 0 otherwise
ja	1 if the tweet is in Japanese, 0 otherwise
ko	1 if the tweet is in Korean, 0 otherwise
sentiment	The computed sentiment of the tweet, in the range [-1, 1]
hashtags	How many hashtags the author used
user_mentions	How many users the author mentioned (i.e. '@-mentioned')
author_followers_count	The number of followers the author has
author_friends_count	The number of friends the author has
author_statuses_count	How many tweets the author has previously tweeted
retweet_count	How many times people retweeted this particular tweet
favorite_count	How many people 'liked' this tweet (i.e. pressed the heart symbol)
ADJ	Adjective count

[4] Here, feature is used in the machine learning sense; that is, a feature is a defining, enumerable characteristic of the dataset that can be individually measured.

ADV	Adverb count
CC	Coordinating conjunction count
CS	Subordinating conjunction count
ET	Count of words identified as “foreign”, (i.e. not of that language)
I	Interjection/exclamation count
NC	Common noun count
NP	Proper noun count
PREF	Prefix count
PRO	Pronoun count
V	Verb count
PUNC	Punctuation count

id	de	fr	ja	ko	sentiment	hashtags	user_mentions	author_followers_count	author_friends_count	author_statuses_count
874048560471224320	0	0	0	1	0.637573	0	0	2624	56	4974
1111472200119259136	0	0	1	0	-0.70122	1	0	602	547	35245
1112428839244828677	0	1	0	0	0.882995	0	0	371	112	15388
1112645166635106304	0	1	0	0	0.741606	0	0	4986	369	2257
1112848387811037184	1	0	0	0	0.331026	1	0	746	362	22239
1112866640910213120	0	0	1	0	0.908263	5	1	10618	1239	5065
1113034277954891776	0	0	1	0	-0.356977	0	0	240	208	1788
1113071079403241473	0	1	0	0	-0.373127	0	0	473	98	6021
1113088785972002818	0	0	0	1	-0.557624	0	0	4099	750	8047
1113241307340783616	0	1	0	0	-0.979918	0	0	460	1305	2676
1113356792094806016	0	0	1	0	-0.607119	0	0	1561892	18935	127146
1113358092702507008	0	1	0	0	-0.542443	0	0	10224	66	595
1113405974369382401	0	0	1	0	-0.929916	0	0	16572	983	11978
1113416205774151680	0	0	0	1	0.80596	0	0	26903	157	34814
1113442548771688449	1	0	0	0	0.333465	6	1	6057	933	5808

retweet_count	favorite_count	ADJ	ADV	CC	CS	ET	I	NC	NP	PREF	PRO	V	PUNC
29	51	5	1	0	0	0	1	16	1	0	1	8	9
16	13	4	2	0	0	0	0	13	0	2	0	21	5
8801	7061	1	0	0	0	0	0	0	0	0	2	2	0
9	102	4	0	1	0	0	0	6	2	0	0	1	2
2	19	1	2	2	0	0	0	6	2	1	7	5	3
2258	883	4	0	0	0	0	0	34	2	0	0	8	6
22334	30236	4	2	0	0	0	0	14	0	1	1	14	9
1975	2510	1	2	0	1	0	0	9	0	0	3	4	4
1600	1386	5	2	0	0	1	0	21	0	0	1	10	7
2	0	5	0	1	0	0	0	8	0	0	0	1	1
1078	1505	3	1	0	0	0	0	16	3	0	0	11	9
5252	12081	2	3	0	0	0	0	3	0	0	4	3	3
30022	70755	4	1	0	0	0	0	10	2	0	1	5	3
1318	1420	3	1	0	0	1	0	2	1	0	0	1	2
2	10	10	2	0	1	0	0	7	0	3	1	4	4