Curses and Swears

Julia Gong, Ben Newman

April 4, 2018

Warning: this write-up contains explicit language

We were very interested in further investigating how swears are represented semantically in word embeddings trained on large datasets. We looked at 2 word embeddings from two sources: CommonCrawl (42B tokens, 1.9M vocab, uncased, 300d vectors) and Twitter (2B tweets, 27B tokens, 1.2M vocab, uncased, 25d, 50d, 100d, and 200d vectors), both found on the glove webpage.

1   Intro

During the class on swearing, we were exposed to the concept that swears do not behave semantically like other words because they serve a more emotion-related than content-related purpose in language. As such, they are used in contexts where their literal meanings do not make sense. As evidence, we were presented with a table of word embeddings for common words and a common swear, and the common words tended to be close to words with similar meanings, but swears seemed to cluster with other common words unrelated to the swear semantically, as below:

3.9   Connotations

Connotations and co-occurrences hypothesis: A word's connotations are reflected in the words that it tends to co-occur with.[5]

Word	Neighbors
speech	free, speak, political, speaking, calling, sort, terms, freedom, called, group
movie	love, watching, family, girl, guys, mind, couple, head, hell, back
video	people, thing, time, playing, part, telling, play, watch, statement, called
mccain	john, obama, barack, campaign, obama's, presidential, candidate, election, senator, race
america	great, terrorists, hope, protect, step, work, nation, lives, tonight, americans
tax	taxes, income, spending, property, cut, rate, pay, measure, economic, lower
sex	life, woman, women, happy, gay, young, thing, love, called, parents
fucking	time, imagine, here's, understand, things, wanted, wrong, hell, stop, stay

Table 1: Connotations extracted from a 90 million corpus of political and gossip blogs.

We decided to try a similar experiment, looking at the top 10 closest words to various swear words, with larger corpora: CommonCrawl and Twitter. First, we investigated the same words that were used in the handout. Next, we moved on to analyzing and classifying swears into two broad categories. Finally, we

looked at how swears might be related to offensive and non-offensive forms of words through an analogy task.

2   Initial Exploration

We began by looking at the same words that were in the class handout in the CommonCrawl corpus embeddings.

CommonCrawl - Chris's words

Word	Neighbors
speech	speeches, speaking, speach, language, voice, words, speak, spoken, remarks, speaks
movie	movies, film, films, starring, dvd, flick, soundtrack, trailer, cinema, picture
video	videos, youtube, clip, clips, hd, audio, footage, pictures, picture, streaming
mccain	romney, obama, hillary, barack, palin, biden, clinton, republicans, gop, huckabee
america	american, nation, europe, states, country, americas, world, americans, united, usa
tax	taxes, taxation, income, taxpayers, revenue, pay, taxpayer, payroll, federal, irs
sex	porn, lesbian, xxx, teen, sexual, gay, nude, porno, anal, sexy
fucking	fuck, ass, fucked, suck, fuckin, pussy, shit, fucks, damn, asshole

As can be seen, there are some differences between the handout closest words and the CommonCrawl closest words, especially when it comes to the words "sex" and "fucking". "Sex" in CommonCrawl seems like a much more vulgar word, which might be expected when considering the entire Internet as a corpus. There are many unsavory figures out there who may make such associations... Similarly, "fucking" in CommonCrawl is also associated with more vulgar, emotionally charged words. Some of them, however, even relate to "fucking" semantically. The cosine similarities for these words in this corpus are shown in the table below:

speech	movie	video	mccain
speeches : 0.65	movies : 0.83	videos : 0.81	romney : 0.85
speaking : 0.64	film : 0.76	youtube : 0.74	obama : 0.83
speach : 0.60	films : 0.72	clip : 0.69	hillary : 0.82
language : 0.60	starring : 0.65	clips : 0.69	barack : 0.78
voice : 0.59	dvd : 0.65	hd : 0.68	palin : 0.77
words : 0.59	flick : 0.64	audio : 0.67	biden : 0.77
speak : 0.59	soundtrack : 0.64	footage : 0.66	clinton : 0.75
spoken : 0.58	trailer : 0.64	pictures : 0.64	republicans : 0.74
remarks : 0.58	cinema : 0.63	picture : 0.64	gop : 0.74
speaks : 0.57	picture : 0.62	streaming : 0.63	huckabee : 0.73

america	tax	sex	fucking
american : 0.74	taxes : 0.87	porn : 0.82	fuck : 0.89
nation : 0.71	taxation : 0.78	lesbian : 0.74	ass : 0.80
europe : 0.71	income : 0.74	xxx : 0.74	fucked : 0.79
states : 0.69	taxpayers : 0.67	teen : 0.73	suck : 0.78
country : 0.69	revenue : 0.66	sexual : 0.73	fuckin : 0.77
americas : 0.69	pay : 0.64	gay : 0.73	pussy : 0.74
world : 0.68	taxpayer : 0.64	nude : 0.72	shit : 0.73
americans : 0.67	payroll : 0.64	porno : 0.71	fucks : 0.73
united : 0.66	federal : 0.64	anal : 0.69	damn : 0.72
usa : 0.66	irs : 0.64	sexy : 0.69	asshole : 0.72

We observe a similar result in the Twitter corpus (using the 100-dimension vectors):

Twitter (100d) - Chris's words

Word	Neighbors
speech	speeches, speaking, letter, interview, during, president, presentation, conference, response, obama
movie	movies, episode, story, trailer, watching, watch, scene, twilight, watched, horror
video	videos, youtube, cover, vídeo, playlist, clip, audio, official, post, trailer
mccain	senator, boehner, congressman, palin, biden, mcconnell, republicans, kerry, christie, cheney
america	american, world, country, mexico, africa, canada, states, us, uk, believe
tax	taxes, debt, budget, pension, insurance, costs, cost, pay, welfare, government
sex	porn, pussy, anal, lesbian, naked, threesome, sexual, horny, boobs, hardcore
fucking	fuckin, freaking, fuck, damn, stupid, fucken, shit, freakin, hell, seriously

"Sex" is similarly vulgar, but "fucking" is less semantically related to the words closest to it compared to the CommonCrawl embeddings, although the emotional content is similar.

In both of these cases, the word "fucking" is associated with other swear words, which isn't the case in the handout. This implies that these other swear words are being used in similar ways to each other, even if their literal meanings differ. The large corpora embeddings, upon an initial review, seem to refute the idea that it doesn't make sense to consider the semantic embeddings of swears because they are used in semantically-inappropriate situations. Instead, they seem to support the idea that swears words are a class of words that are used contrary to their literal semantics. To further explore this idea, we wanted to investigate representations of other swear words in these two corpora.

3   Swear Words

We started by looking at a variety of different swear words:

Word	Neighbors
fuck	fucking, ass, fucked, suck, bitch, pussy, shit, fucks, slut, horny
damn	freaking, thats, darn, shit, yeah, stupid, cuz, freakin, lol, fuckin
bitch	bitches, fuck, ass, whore, slut, fucking, shit, fucker, suck, asshole
hell	shit, damn, heck, yeah, crap, gonna, thing, heaven, thats, nothing
shit	crap, fuckin, damn, cuz, thats, lol, dude, fuck, stupid, sht

Some of the words ("damn", "hell", and "shit") seem to be associated with words that aren't swears (such as "yeah", "thats", and "gonna"). Those are consistent with the handout's associations and show that these words may be used in different emotional contexts from other swears. "Bitch" and "fuck", however, seem to be more associated with not only more intense swears but also more semantically relevant words. In the next section we try to uncover why this might be by investigating two overly general classes of swears: mild and intense swears.

4   Classifying Swears

4.1   Mild Swears

We decided to refer to the swears that tended to co-occur with more mild vulgarities or non-offensive emotional words mild swears.

CommonCrawl - Mild Swears

Word	Neighbors
dang	darn, damn, gosh, freaking, soo, freakin, friggin, darned, yah, god-damn
darn	darned, damn, freaking, dang, soooo, freakin, friggin, pretty, awfully, sooo
heck	yeah, hell, damn, maybe, anyway, guess, thing, think, thats, wondering
gosh	geez, jeez, gawd, omg, ohh, ohhh, ahhh, ahh, darn, ugh
jeez	geez, geeze, jeeze, sheesh, damnit, gosh, ummm, dammit, c'mon, woah
hella	soooo, sooo, sooooo, fuckin, freakin, soooooo, gettin, damn, kinda, soo
freaking	freakin, friggin, damn, effing, goddamn, fcking, fuckin, darn, frickin, fking

Twitter (100D) - Mild Swears

Word	Neighbors
dang	damn, dam, nahh, soo, nah, huh, idk, gah, noh, lol
darn	friggin, freakin, effing, freaking, fricken, gosh, goddamn, effin, frickin, fucken
heck	hell, wth, jeez, gosh, dammit, woah, damn, kidding, crap, yeah
gosh	omfg, omg, ohmygod, ugh, srsly, jeez, omgg, freaking, dammit, seriously
jeez	geez, sheesh, gosh, geeze, jeeez, geesh, damnit, heck, dammit, woah
hella	mad, shits, kinda, gettin, lowkey, fuckin, damn, dumb, slick, alot
freaking	freakin, fricken, effing, fucking, fucken, friggin, frickin, seriously, gosh, fricking

Many of these words are associated with emotional outbursts, both positive and negative, but are not necessarily offensive. Others, specifically those associated with 'heck' in CommonCrawl, aren't even really outbursts and are just common words. These associates better support the theory proposed in the class handout.

Side note: the "hella" entry for both has some really interesting word associations. 'sooo' (with some number of 'o's makes sense because 'hella' colloqially means 'a lot'. The Twitter corpus shows 'hella' to be associated with 'lowkey' and 'kinda', two words that are not semantically associated with the concept of 'a lot'. They actually tend to suggest the opposite. It would be interesting to investigate if these words are being used in this context because of politeness, humbleness, hedging, or an attempt to come off as more personable or approachable.

4.2   Intense Swears

Of course there swears that are much less mild as well. These seem to be more likely to be used in ways that directly correlate to their (usually sexual) semantics. This might be because these swears are used relatively less frequently, so if an utterance uses one in a situation where the literal meaning is not relevant, it might be more marked. This rarity of usage also gives them their power.

CommonCrawl - Intense Swears

Word	Neighbors
damn	freaking, thats, darn, shit, yeah, stupid, cuz, freakin, lol, fuckin
shit	crap, fuckin, damn, cuz, thats, lol, dude, fuck, stupid, sht
fuck	fucking, ass, fucked, suck, bitch, pussy, shit, fucks, slut, horny
bitch	bitches, fuck, ass, whore, slut, fucking, shit, fucker, suck, asshole
cunt	pussy, twat, asshole, cunts, cock, ass, fucking, clit, slut, fuck
dick	cock, dicks, sucking, suck, ass, pussy, fucking, fuck, cocks, cunt

Twitter (100D) - Mild Swears

Word	Neighbors
damn	shit, fuckin, fuck, smh, hell, fucking, lmao, that, why, thats
shit	fuck, damn, that, nigga, hell, thats, bitch, like, smh, lmao
fuck	shit, hell, bitch, fucking, damn, wtf, fuckin, n't, that, nigga
bitch	nigga, fuck, ass, bitches, hoe, shit, girl, fuckin, damn, cuz
cunt	twat, prick, bastard, cunts, wanker, dickhead, faggot, fucker, asshole, bellend
dick	suck, ass, cock, pussy, bitch, dicks, nigga, penis, asshole, fuck

While all of these words are words that you would scold your child for using, there are some interesting differences between their associations. For "damn" and "shit" there is almost no overlap between the dictionary definition and the way it is used. Like the more mild swears, they tend to be used similarly to exclamatory markers like 'crap' and 'lol' as well as other common words like 'cuz' and 'that'. What sets these words apart from the mild swears, however, is that the other swear words that they are associated with are also more intense. This makes sense, though, as it would sound strange to mix mild and intense swears (as in 'The darned fucking dog'), and usage of these swears in the same place in a sentence would convey very different sentiments (as in 'The fucking dog' when compared to 'The darned dog').

For "fuck" and "bitch", however, while some of the most similar words are merely other intense swears, some seem to be a little more semantically related. For instance, "bitch" is associated with "whore", "slut", and "girl", which all share the gender typically being expressed. "Fuck" is semantically related to quite a few of its associated words, especially in the CommonCrawl embeddings.

"Cunt" and "dick", however, are even more associated with words that share their semantics − particularly "cunt". These associated words still tend to be quite vulgar, so they all express a similar emotion, but they *also* have similar meanings.

These words seem to exhibit a correlation between the intensity of a swear word and the contexts that it can be used in. In addition to expressing intense emotion, it seems like relative mildness of "damn" and "shit" compared to "cunt" and "dick" can roughly be seen in the number of a) strong and b) semantically related synonyms it is associated with (with the more shared synonyms signaling markedness in daily conversation and thus relatively higher strength).

5 Analogies

Finally, we tried to isolate the strong emotional component of a swear's embedding using an analogy task. These analogies were of the form "fuck" is to "sex" as "hell" is to ___. To do this, we found the ten closest embeddings to the vector specified by adding the difference between a swear and non-swear form of a word to a non-offensive word (for example: "fuck" − "sex" + "heck", or "douche" − "fuck" + "nice"). Below are some of the pairs we looked at:

Twitter (100d) - Analogies

Positives	Negatives	Neighbors
fuck, **heck**	**sex**	hell, dammit, eff, fuuuuck, damn, fuq, wtf, maaan, damnit, fuuuck
fuck, **poop**	**sex**	fucker, poo, crap, outta, shit, shut, fucking, piss, fuckin, dumbass
fuck, **darn**	**sex**	friggin, dammit, damnit, freakin, effing, goddamn, effin, fucken, freaking, damn

The words in the 'positives' column represent the vectors we're adding and the ones in the 'negatives' column represent the vectors we're subtracting. The 'neighbors' are the 10 closest words to that vector sum. The analogies somewhat support what we were looking at, with "hell" being the top word associated with "heck", for instance. However, it really seems like some swears, like "shit", are used in contexts so far from their literal semantic meanings that they are not just some emotional form of their non-offensive counterparts. For example, "shit" is not used simply as a more emotional word for "poop", as we can see it doesn't have a high association to "poop" in the table above.

If our observations about more intense swears hold, however, we should expect the more intense swears to hold more semantic content and thus to perform better on the analogy task:

Positives	Negatives	Neighbors
cunt, **penis**	**vagina**	prick, twat, bastard, dickhead, wanker, faggot, bellend, lad, fag, cunts
dick, **vagina**	**penis**	pussy, bitch, ass, mouth, whore, nasty, suck, fuck, wet, shut

Interestingly, this small subset of intense swears perform just as well as the perhaps less intense ones. This may be the beginning of evidence that more intense swears are not necessarily used in more semantically appropriate circumstances.

Another option to consider is that common swears like "fuck" and "shit" are so detached from their semantics that their word vectors act as pure emotional components; namely, a strong negative emotion with a dominating semantic force. In the above examples, this would mean we wouldn't have to subtract the vector for "sex" because "fuck" is often used in such different contexts from "sex" anyway (this kind of disagrees with our analysis in part 5 saying that "fuck" is somewhat semantically related to its closest neighbors). It is easy enough to look at some examples of this.

Twitter (100d) - Pure Emotion Swears

Positives	Negatives	Neighbors
butt, **fuck**	*NONE*	ass, dick, bitch, shit, damn, like, hell, head, fucked, shut
twat, **fuck**	*NONE*	cunt, asshole, fucking, fucker, shut, stupid, bastard, shit, bitch, prick
darn, **fuck**	*NONE*	damn, fucking, fuckin, stupid, hell, freaking, ugh, heck, gosh, shit
heck, **fuck**	*NONE*	hell, damn, wtf, shit, why, yeah, dude, fucking, shut, fuckin
penis, **fuck**	*NONE*	dick, shit, like, bitch, fucking, that, ass, suck, hell, damn
linguistics, **fuck**	*NONE*	sociology, geography, physics, biology, psychology, science, economics, studying, shit, stupid

This looks promising, and very similar results were found when using "shit" instead, because it turns out that in this corpus "shit" and "fuck" are very similar vectors (with a cosine similarity of 0.918). This was investigated in other corpora as well, but the results did not line up as well:

CommonCrawl - Pure Emotion Swears

Positives	Negatives	Neighbors
butt, **fuck**	*NONE*	ass, fucking, pussy, fucked, cock, suck, bitch, booty, cunt, porn
twat, **fuck**	*NONE*	cunt, pussy, fucking, asshole, ass, fucked, fucks, slut, cock, horny
darn, **fuck**	*NONE*	damn, fucking, shit, fuckin, thats, yeah, freaking, stupid, dude, lol
heck, **fuck**	*NONE*	shit, yeah, fucking, damn, hell, thats, dude, cuz, lol, guys
penis, **fuck**	*NONE*	fucking, cock, pussy, ass, suck, dick, cunt, sucking, sex, fucked

Similarly, we can also observe this phenomenon of swear words serving as emotional components of a word when subtracting them from very strong words that have weaker counterparts, as seen in the table below.

Positives	Negatives	Neighbors
dumbfuck	**fuck**	dimwit, stupidhead, *omitted phrases from other languages*, sycophant, fucknut, dumbdumb, slutbucket
vagina	**fuck**	vulva, dentata, buttocks, vigina, *omitted phrases from other languages*
dick	**fuck**	flaccid, bootyhole, aftera, clit, squirter, uncircumcised, fingertip, atlayan, *omitted phrases from other languages*

Initially, it appears that the results are largely what we expect: it appears that the words most associated with "dumbfuck" after the purely emotional component of the words given by "fuck" are removed are words that have the same semantic meaning as the original stronger word, but with a weaker force, such as "dimwit", "dumbdumb", and "stupidhead".

However, upon closer examination, we also see that this is not exactly the case for "vagina" and "dick", which have inherently sexual associations. It is particularly interesting to note that the words associated with these two words after "fuck" is subtracted from the word vector are more on the side of *euphemisms* for these more vulgar, strong terms. Thus, we also propose that "fuck" may not only add strength of emotion, but also perhaps directness or a sense of unfiltered vulgarity to words.

It seems like the fact that we're adding emotion-full swear words overpowers a lot of the semantics that the non-offensive words contribute, but it was worth investigating this idea of certain swears' embeddings representing pure emotional components of a word.

6 Conclusion

From the large corpora we examined, it appears that semantic embeddings of swears do give information about their use. We found that swears tend to be used in similar contexts as other swears and emotional words. We also observed that milder swears tend to be associated with other milder swears and more intense swears tend to associate with other more intense swears. Additionally, it appears that the more intense a swear is, the more likely its vulgar usage actually refers to something similar to its semantic meaning, but more analysis would be needed on this front, especially given the analogy task we investigated.

We also were able to discover some evidence for the usage of strong swear words like "fuck" to convey purely emotional states rather than semantically encoded meanings through 'adding' and 'subtracting' the strong swears from strong or weak words.

For future research, it would be interesting to see how these word associations changed over time (for example, if "damn" was closer to other hell-related words in the 1800's).