Expert Systems With Applications 198 (2022) 116862

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

One emoji, many meanings: A corpus for the prediction and disambiguation of emoji sense

Matthew Shardlow [a,*], Luciano Gerber [a], Raheel Nawaz [b]

[a] Department of Computing and Mathematics, Manchester Metropolitan University, United Kingdom

[b] Operations, Technology, Events and Hospitality Management Department, Manchester Metropolitan University, United Kingdom

ARTICLE INFO

Keywords:
Text mining
Semantic networks
Emoji

ABSTRACT

In this work, we uncover a hidden linguistic property of emoji, namely that they are polysemous and can be used to form a semantic network of emoji meanings. Our key contributions to this direction of study are as follows: (1) We have developed a new corpus to help in the task of emoji sense prediction. This corpus contains tweets with single emojis, where each emoji has been labelled with an appropriate sense identifier from WordNet. (2) Experiments, which demonstrate that it is possible to predict the sense of an emoji using our corpus to a reasonable level of accuracy. We are able to report an average path-similarity score of 0.4146 for our best emoji sense prediction algorithm. (3) We further show that emoji sense is a useful feature in the emoji prediction task, where we report an accuracy of 58.8816 and macro-F1 score of 46.6640, beating reasonable baselines in this task. Our work demonstrates that importance of considering the meaning behind emoji, rather than ignoring them, or simply treating them as extra wordforms.

1. Introduction

Emoji are used commonly in informal communications such as private messages or social media (Hurlburt, 2018). They are used to indicate and reinforce the author's intended meaning or sentiment as attached to a text. Emoji are not a sub-language, but they do bear semantics — typically functioning as semantic interjections (Na'aman et al., 2017). Any natural language processing system working with informal text that ignores emoji is missing out on a vital source of author intent.

Emoji are typically treated as monosemous units. However, this is clearly untrue to anyone who is familiar with emoji. One emoji may be used in multiple contexts (Donato & Paggio, 2017), e.g., the fire emoji may indicate physical attractiveness, heat, actual fire, etc. Similarly, multiple emoji may be used to mean the same thing. E.g., the heart emoji, heart eyes, or two-hearts emoji may be used interchangeably in circumstances where they are used to indicate love.

This structure of one lexeme (i.e., a basic unit of meaning, such as a word or an emoji) having multiple meanings and multiple lexemes being used interchangeably is not unfamiliar, what we are describing here is the concept of a WordNet (Miller, 1995). The same semantic structure which is commonly used and understood for words can be applied to emoji. By developing a semantic network for emoji, we can better understand the relationships in meaning between emoji. We can also perform natural language processing techniques at the level of emoji meaning, rather than just looking at the form of the emoji itself.

To do this we must treat emoji as lexical units in their own right. Whilst emoji do not function in the same way as words (Na'aman et al., 2017), there are similarities in the way that they can be approached and recent research has shown that they can be categorised semantically in the same way as words (Eisner et al., 2016). The semantic categorisation that we are proposing goes beyond previous attempts as we are suggesting the creation of a semantic network of emoji, rather than merely developing linguistic tools to enable the usage of emoji in NLP (Illendula & Yedulla, 2018).

In this paper, we have taken the idea of emoji semantics and made a first attempt at developing a purely data-driven semantic network for emoji. Whilst other emoji semantics networks do exist, we discuss their deficiencies compared to our approach in Section 2. Our key contributions are as follows:

1. To aid in our analysis, we collect a corpus of 721,505 tweets in Section 3, where each tweet contains just one emoji.
2. We annotate a partition of our corpus with sense labels from WordNet and report on the features of the annotated and unannotated portions of our corpus in Section 4, demonstrating the polysemous nature of emojis that we have posited in this introduction.

* Corresponding author.
E-mail addresses: M.Shardlow@mmu.ac.uk (M. Shardlow), L.Gerber@mmu.ac.uk (L. Gerber), R.Nawaz@mmu.ac.uk (R. Nawaz).

https://doi.org/10.1016/j.eswa.2022.116862
Received 30 April 2021; Received in revised form 4 November 2021; Accepted 8 March 2022
Available online 18 March 2022

3. We describe a methodology for unsupervised prediction of emoji senses using a modified Lesk algorithm based on embeddings in Section 5.
4. The emoji sense is shown to be a useful feature for the related task of emoji prediction in Section 5.5.
5. Finally, we describe an algorithm that can be used to create a semantic network of emojis and apply this algorithm to the unannotated portion of our dataset. We describe the features of the resultant emoji network, focusing on 8 emoji that we have studied in the annotated portion of our corpus.

2. Background

Word Sense Disambiguation (WSD) is a well known task in Natural Language Processing, where the aim is to take a word in context w from some vocabulary W and a set of senses S defined as $s_1 \dots s_n$ and assign w to some s_i. Further words w_i in context may also be assigned to the senses in S. Eventually, a semantic network (or WordNet) is formed, mapping W to S. Synsets can be defined as the set of elements of W that map to one s_i. Each element in w_i has a set of senses defined as those elements in S to which it maps.

The original effort to create such a machine-readable semantic network is Princeton Wordnet (Miller, 1995) (often simply referred to as WordNet), which was developed by manually analysing word senses in the Brown corpus (Francis & Kucera, 1979). WordNet is freely available as a semantic resource, and is integrated into many modern NLP APIs. WordNet goes further than the semantic network as described above, also defining relationships such as antonymy, meronymy, and hypernymy (Fellbaum, 2010). Further developments in this area have seen the creation of semantic networks for languages other than English (such as French Sagot & Fišer, 2008, or Chinese Wang & Bond, 2013), multi-lingual semantic networks (Bond & Foster, 2013; Fellbaum & Vossen, 2012) (wherein words from multiple languages all map to a common set of senses S) and semantic networks incorporating domain information (Navigli & Ponzetto, 2010).

In our work, we are aiming to create the same type of network, focusing on emoji. Our network comprises of a set of Emoji in context E which are mapped to a set of senses S derived from WordNet. In the remainder of this short literature review we will first cover the area of WSD, giving treatment to the shared tasks and recent advancements. We will then cover the treatment of Emoji both outside of the field of NLP and within the field of NLP, particularly focusing on similar efforts to categorise the semantics of emoji. Finally, we will cover other resources that aim to categorise emoji, showing where prior efforts are deficient and motivating our novel approach to developing a semantic network for emoji.

2.1. Word sense disambiguation

We have defined the task of WSD above as it is pivotal to our intended goal of understanding the various meanings of Emoji. A classic algorithm used in WSD is the LESK algorithm (Lesk, 1986). In this algorithm, word overlap between a target's context and the gloss of a wordnet synset is analysed to determine whether any content words may help determine the sense of a word. The field of WSD has been advanced through several shared tasks over the years as covered by the popular survey article from Navigli (2009). These shared tasks began with the SensEval (now SemEval) workshop series (Edmonds & Cotton, 2001; Kilgarriff & Rosenzweig, 2000; Mihalcea et al., 2004a). Through these shared tasks, new datasets on the prediction of word senses were developed and released for the community. Work to improve system's performance on these shared tasks led to state of the art systems for WSD that used a variety of methods (Mihalcea et al., 2004b), including unsupervised (Niu et al., 2004; Ramakrishnan et al., 2004), and supervised learning (Grozea, 2004; Strapparava et al., 2004). Unsupervised approaches tended to use strategies such as dictionary matching (Ramakrishnan et al., 2004) with sense definitions and contexts and clustering of instances into sense groups (Niu et al., 2004). Supervised approaches on the other hand use machine learning approaches such as Naïve Bayes (Grozea, 2004) and SVMs (Strapparava et al., 2004) in order to learn patterns that predict word senses.

GlossBert (Huang et al., 2019) bridges the gap between traditional manually curated sense inventories and modern deep-learning approaches to lexical semantics. The authors of GlossBert combine context and glosses from WordNet and develop new models using BERT that are fine-tuned to achieve state of the art results on word sense disambiguation tasks. Certainly, the future of WSD lies at the intersection of deep learning and human knowledge.

Recently, deep learning for NLP (or DeepNLP) has provided new insights into the field of WSD. The ELMo language model was developed to produce contextual embeddings for every word in a sentence. Each word has its own embedding, which is influenced by the context in which that word is found. This means that two tokens with identical wordforms will only have similar embeddings if their contextual usages (and hence meanings) are the same. Similarly, tokens used in similar contexts (hence indicating synonymy) will also exhibit similar embeddings. This property is not only true of the ELMo model, but also an emergent property of transformer architecture language models such as BERT (Devlin et al., 2019), XLNet (Yang et al., 2019) and the GPT-family (Budzianowski & Vulic, 2019; Floridi & Chiriatti, 2020).

2.2. Emoji in the literature

Emoji have formally existed since 2009 as part of the Unicode standard 5.2, however the usage of pictographics dates back to prehistory (Hurlburt, 2018). Modern Emoji are ubiquitous, thanks to their integration into smartphone keyboards (Lu et al., 2016). There is significant variation in the usage of emoji between different cultural groups across the world (Ljubešić & Fišer, 2016; Lu et al., 2016).

Emoji are typically used in one of 3 ways. Either as a content word: "I $\langle Heart \rangle$ New York!!", or as a function word in a redundant expression, repeating and intensifying information in the text: "I love New York!! $\langle Heart \rangle$" or in a non-redundant manner adding additional information to a text: "New York!! $\langle Heart \rangle$". Where $\langle Heart \rangle$ is replaced with the relevant emoji in each case (Donato & Paggio, 2017; Na'aman et al., 2017).

Although there is a significant body of work covering emoji as demonstrated above, the treatment of emoji with Natural Language Processing techniques is limited to relatively few studies. We have surveyed these below.

Similarly to the WSD task, emoji prediction has been advanced through a series of shared tasks as part of the annual SemEval workshop. Emoji prediction is somewhat different to our task of emoji disambiguation, however it warrants investigation here as an interesting field of study.

A similar task to that of emoji prediction is the one of predicting the similarity of emojis. In this task, 2 emojis may be considered similar if they are typically used in a similar manner (e.g., to express humour), or may be considered different if they are typically used to express something different. The EmoSim508 dataset (Wijeratne et al., 2017b) contains 508 such pairs of emojis with human assigned scores. The authors provide machine learning techniques which are able to accurately predict the similarity and dissimilarity of their emoji pairs.

A popular usage of emojis is to clarify and intensify the sentiment expressed in a sentence. Someone might use an angry face to further express their anger, or a laughing face to express their joy on top of the text they have already written. These representations of sentiment cannot be overlooked, especially in social media text. Indeed incorporating the emoji into sentiment analysis leads to state of the art performance (Felbo et al., 2017).

With the rise of Deep Learning as applied to Natural Language Processing, there has been a shift towards incorporating emoji into deep learning architectures for NLP. One such approach, Emoji2Vec (Eisner et al., 2016), learns the embeddings for each emoji by looking at the description of that emoji in the Unicode standard. Clearly, this is problematic as each emoji is limited to a maximum of one embedding. Further, the description of an emoji may not be representative of its real usage. Work does exist to create emoji embeddings from real contextual usages (Guibon et al., 2018), overcoming the issue stated previously. Emoji embeddings may also be learnt by looking at emoji co-occurrences and developing an emoji co-occurrence network, which can be turned into a matrix giving embeddings through dimensionality reduction (Illendula & Yedulla, 2018). The state of the art model for deep learning with emoji is the aptly named DeepMoji (Felbo et al., 2017). In this work emoji embeddings are learnt as a by-product of an emoji prediction task on a large and varied dataset. These resulting embeddings are useful for a variety tasks, including sentiment analysis. Embeddings of this sort are useful for downstream tasks such as sarcasm detection (Felbo et al., 2017) or emotion recognition (Ahanin & Ismail, 2020).

Emoji were integrated into the Bert language model (Devlin et al., 2019) giving state of the art performance on a question answering task (Delobelle & Berendt, 2019). Bert is a state of the art pre-trained language model built on the transformer architecture (Vaswani et al., 2017) which is capable of learning syntactic and semantic relationships between words. Bert provides contextual embeddings for each token, including each emoji in a sentence, allowing emoji to be treated as other tokens and incorporated in DeepNLP pipelines as textual features.

2.3. Other categorisations of emoji

There are three outstanding resources that warrant discussion as categorisations of Emoji. These are Emojipedia, the Emoji Dictionary and EmojiNet. Each is developed from a different perspective and is popular in its own right as an Emoji resource. However, none of these resources meet the criteria of a semantic network as discussed earlier. We have described each resource below and outlined their shortcoming for our purposes.

Emojipedia[1] is an online website which has definitions provided for each emoji by professional lexicographers. The definitions are provided as free text and are built on the basis of the lexicographer's understanding of that emoji as well as factors such as contextual usage. According to the Emojipedia website, all definitions have been provided by 3 lexicographers[2] Whilst Emojipedia is useful for lay users of emoji who wish to better understand their meanings, it is not useful for machine ingestion as it only provides a free text component to the data, rather than any structured network (semantic or otherwise). There are useful features which could be of use for natural language processing such as Emoji categories, alternative names and links to related emoji, however there is no academic literature currently making use of the Emojipedia data of which we are aware.

The Emoji dictionary[3] is an online wiki site which crowd-sources definitions of emojis from its user base. Whilst this approach allows for quick creation of a resource covering the meanings of a wide number of emoji it has a clear limitations in that any new definition of an emoji is accepted. This leads to noisy data in which each emoji has hundreds of submissions. The definitions are provided as free text, rather than as semantic categorisations, meaning that links between emojis (i.e., those with the same meaning) are likely to be missed. To use the emoji dictionary well, any system would certainly need to perform a significant amount of data cleaning to remove spurious or offensive submissions, as well as to standardise the language used in the submissions.

Emoji	Emoji Name
❤️	Heart
💕	Two Hearts
😩	Weary Face
😘	Face Blowing Kiss
😊	Smiling Face with Smiling Eyes
😍	Heart Eyes
😒	Unamused Face
😭	Loudly Crying
♥	Black Heart Suit

Fig. 1. The nine emoji selected in our study.

EmojiNet (Wijeratne et al., 2017a) is an academic project focused on the development of a machine readable resource of emojis. This is the most similar existing work to ours and deserves special treatment as part of this literature review. Briefly, in their work, (Wijeratne et al., 2017a) propose to develop a network of emojis with associated senses and semantic links between these emoji. They do this by mining the aforementioned Emoji Dictionary and linking concepts found therein for each emoji to BabelNet synsets. This work is clearly a step in the right direction, however suffers from the reliance on external resources. The Emoji Dictionary is not curated and there is no guarantee of correctness or suitability in the suggestions provided there. Further, if a definition term provided in the Emoji Dictionary is not available in BabelNet then this sense cannot be integrated. It should be noted that the linking of a text-based term to a BabelNet synset is non-trivial as any term may link to more than one synset, and so a WSD algorithm must be employed to resolve this, leading to the potential introduction of errors in the EmojiNet corpus.

Our work differs from that of EmojiNet in that we do not rely on external resources such as BabelNet and the Emoji Dictionary, but instead we produce a purely data driven methodology for identifying the senses present in Emoji. This is less prone to the potential pitfalls outlined above.

3. Data collection and annotation

We initially collected a corpus of emoji-bearing tweets using the Twitter Developers API between July and September 2018. The tweets we collected were filtered to ensure that each tweet only contained a single Emoji (negating compositional effects on an Emoji's meaning). We also applied a series of hand-crafted rules to ensure that the tweets came from real users of Twitter as opposed to bots. For example, we removed tweets that occurred with the same text very frequently (typically being advertisements) and removed users who tweeted with very high frequency over a short period (typically being spammers). This led to an initial corpus of 721,505 tweets spread across 239 emoji.

This left us with a large corpus of the type of emoji in use that we are interested in, however no information about the intended meaning of the emoji was present in each tweet. We decided therefore to perform an annotation round that would allow us to better understand the sense of the emojis being used. We selected the nine Emoji that occurred most frequently in our corpus and annotated a sample of these for word sense. The emoji we selected are shown in Fig. 1.

To ensure that the sample we selected was sufficiently varied, we first clustered the tweets for each emoji using the Infersent embedding of the emoji text. InferSent is a publicly available system that takes

[1] https://emojipedia.org/
[2] https://emojipedia.org/meanings
[3] https://emojidictionary.emojifoundation.com/

Table 1
Statistics on the corpus after the annotation round. We report raw values as well as those after reducing the number of senses in our corpus. For each view of the senses we report the number of senses for each emoji, as well as the top sense associated with each emoji. After the reduction, the most common sense changes in some instances, leading to the reported sense differing between raw and reduced. The original sense from the raw data is still preserved as a less frequent sense in the list of senses associated with that emoji.

	Raw		Reduced	
Emoji name	Senses	Top sense	Senses	Top sense
Black heart suit	57	Wishful	20	Well-wishing
Two hearts	47	Insecurity	6	Eager
Smiling face with smiling eyes	57	Positivity	16	Weary
Weary face	55	Enchantment	10	Upset
Face blowing kiss	43	Attractiveness	9	Injustice
Loudly crying	78	Displeased	35	Achievement
Heart	61	Felicitation	24	Love
Unamused face	40	Wishful	13	Endearment
Heart eyes	51	Plead	19	Disbelief

a sentence and provides a vector of numbers which embeds contextual information representing the sentence's meaning (Conneau et al., 2017). We created 8 clusters and selected at random 25 tweets from each cluster to give 200 tweets of varied senses for each emoji — giving a total corpus of 1800 tweets.

We then annotated each of the tweets using crowdsourcing via Amazon's Mechanical Turk.[4] Each tweet was given to 5 crowd workers who were asked to identify a Wordnet synset (using a web interface) that corresponded to the intended meaning of the emoji. We did not require annotators to have any specific linguistic qualification beyond speaking English. We regularly checked the annotator pool to ensure that annotators were doing the task. We rejected any annotators that did not perform the task (i.e., leaving the answers blank) to ensure annotation quality. We provided an introduction to the task for annotators to read through which is included as supplementary material to this paper. A research assistant was employed to process the results of this round of annotations. The research assistant did not provide new annotations, but rather selected the most appropriate senses for each tweet based on the annotations given during the crowdsouring.

4. Corpus statistics

In this section we present statistics on our annotated corpus as well as some discussion of its practical uses. The statistics for the first version of our corpus are presented in Table 1. We also produced a second version of our corpus in which the number of senses has been reduced by combining senses which share a common hypernym. The statistics for the reduced corpus are also presented in Table 1.

It is clear from the data we have presented that our emoji are highly polysemous. In the raw portion of the data, each of the emoji had between 40 (Unamused Face) and 78 (Loudly Crying) distinct usages. This is based on a sample of 200 of each emoji and so may be an under representation of the true number of senses for each. Emoji are used widely without reference to a universal dictionary of meaning, so it is an intuitive result to demonstrate the wide variety of meanings that one emoji may take. We expect this to also be true for other emoji beyond the nine studied here.

We have also included a second set of results in the last two columns under the heading 'reduced'. To create these results we examined the labels that had been given to our emoji and checked Wordnet to see if there were any hypernyms that covered 2 labels. This allowed us to combine labels with similar meanings, e.g., the senses 'Devotion' and 'Loyalty' may have the hypernym 'Love'. We ensured that hypernyms were at least 3 nodes below the root (i.e., above a depth of 3 nodes) in the tree to prevent the referent concepts from being too abstract as the top of the WordNet tree contains categories that are unsuitable for our purposes (e.g., object, trait, etc.)

The results for our reduced version of our dataset show that fewer senses are given for each of our emoji. The reduction in the number of senses varies between emoji with 'Two Hearts' going from 47 senses down to 6 (an 87% reduction) and 'Loudly Crying' going from 78 to 35 senses (a 55% reduction). This indicates that the annotated senses of some emoji were much closer than in other cases, allowing a greater reduction to be given. The emoji which continue to have a wide number of senses are more likely to be truly polysemous. The reduced dataset still has a median of 16 senses which is more than is typically found in lexical semantic studies. This is again unsurprising as emoji are used without a fixed point of reference and may have different meanings in different groups and subcultures.

For both the raw and reduced sections of our dataset we have included a 'top sense' column, indicating the most frequent sense that occurred for each emoji to give the reader an idea of the types of annotations our emoji were given. Emoji are typically used to express emotions or feelings, and this is reflected in the senses given (e.g., 'positivity', 'wishful', etc.). The reduced senses have different top labels as the labels given are constructed from the hypernyms of the original labels as described above. These labels are still sufficiently descriptive to capture the meanings of each of the emoji.

5. Predicting emoji senses

In the previous two Sections we have described our corpus collection methodology and the composition of the resulting corpus. Whilst the corpus is useful for giving insights into the nature of Emoji by itself, we are also interested in the task of automatically predicting the sense of an emoji. Whilst our corpus is not particularly large, due to the financial constraints on annotating emoji with crowd-workers, we have overcome this limitation to present results on predicting emoji senses in our corpus. To this end, we take an unsupervised methodology in line with our small dataset and the wide number of senses that are covered in our corpus.

5.1. Features

To predict emoji sense We adapt the GlossBert methodology (Huang et al., 2019) to work for emoji sense prediction. GlossBert works by taking the Bert Embedding of a word and comparing it to the embedding for a set of WordNet Glosses. The most similar embedding is the sense that is assigned to a word. Although this algorithm seems simplistic, it is rooted in the historic Lesk baseline WSD algorithm, where content words in a gloss are compared to the context of a given word. We adapt this methodology to work for our emoji by looking at the infersent embedding of the emoji's context (i.e., the tweet in which it appears) to give a text embedding and comparing this via cosine similarity to the gloss embedding for each of the senses that have been identified as part of our annotation process. The most similar gloss is considered to be the correct sense for the given emoji. Our algorithm runs in time complexity $\mathcal{O}(N \times M)$ where N is the number of tweets and M is the number of glosses.

Our methodology is visualised in Fig. 2, where we show a tweet bearing an emoji being converted into an embedding via the InferSent methodology and then being compared to the glosses of a number of senses taken from WordNet. The most appropriate sense is the one with the most similar InferSent embedding to the original text. In our case the original tweet was about loving New York and the sense "Object of warm affection" is selected over other less appropriate senses such as "Feeling of sexual desire".

We further extend our text embeddings by incorporating an embedding for the emoji itself (using emoji embeddings from FastText Bojanowski et al., 2017) and by incorporating a text string representing

[4] https://www.mturk.com/

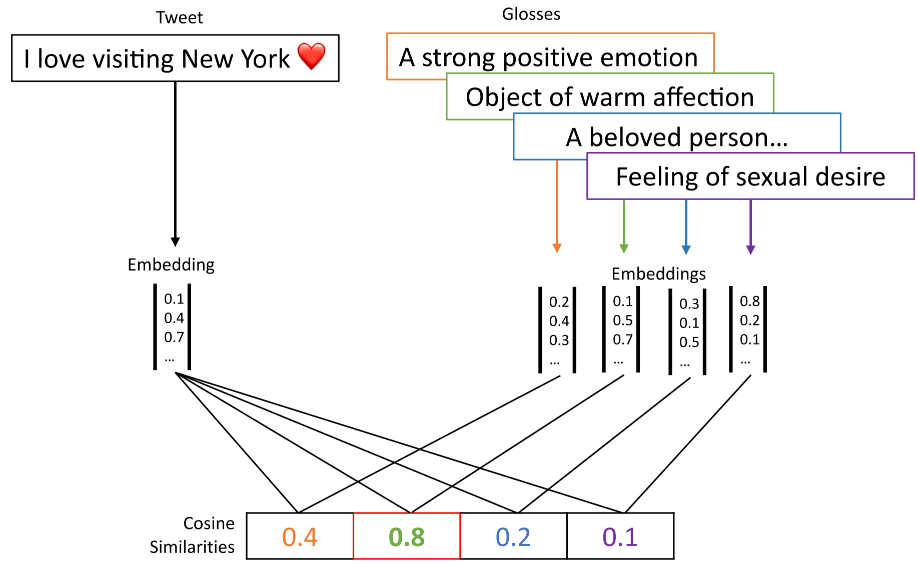

Fig. 2. The emoji disambiguation process.

the emoji. Each of these strategies allows us to incorporate further information into the embedding that is present in the tweet, but would not be captured by conventional means. We also tried incorporating the emoji embedding and emoji string embedding concurrently, but this did not lead to further improvements.

We further extended our gloss embeddings by incorporating hypernym and hyponym embeddings. To do this we concatenate all tokens from the glosses of either the direct hypernyms or direct hyponyms to the original gloss and treat this as a new gloss for which we calculate a new embedding using infersent. We tried incorporating both hypernyms and hyponyms concurrently, but this did not yield higher performance than the analyses we have already reported.

Table 2
Results on predicting emoji senses using our unsupervised and semi-supervised approaches.

Text embedding	Gloss embedding	sim-all	sim-reduced
tweet	gloss	0.1014	0.1321
tweet	gloss + hypernyms	0.0997	0.1175
tweet	gloss + hyponyms	0.1217	0.3390
tweet + emoji	gloss	0.1019	0.1409
tweet + emoji	gloss + hypernyms	0.0987	0.1288
tweet + emoji	gloss + hyponyms	**0.1137**	**0.4146**
tweet + emoji-String	gloss	0.0983	0.1324
tweet + emoji-String	gloss + hypernyms	0.1026	0.1158
tweet + emoji-String	gloss + hyponyms	0.1301	0.3366
Most frequent sense		**0.2256**	**0.6461**

5.2. Evaluation methodology

We initially considered evaluating our approaches using direct accuracy as a metric. However, in our initial experiments, we found this to be overly punitive. In our task, there is some degree of subjectivity and a wrong answer may be close in meaning to the real answer, yet would still be given a score of zero in direct accuracy. We decided to use a custom accuracy metric based on the WordNet path similarity.

Path similarity is defined as the number of steps needed to be taken in WordNet's taxonomy (using hypernym/hyponym relations) to get from one sense to another. This is transformed into a 0–1 range by adding 1 and taking the inverse, such that a distance of 0 steps would yield a path similarity of $\frac{1}{0+1} = 1$, a distance of 1 step would yield: $\frac{1}{1+1} = 0.5$, a distance of 2 steps would yield: $\frac{1}{2+1} = 0.\dot{3}$ and so on. We take the mean average of these path similarity values to give an average path similarity score. This has the advantage of scoring the same for positive cases as simple accuracy would, but also positively rewarding cases where the systems predicts a similar, but not quite right sense.

5.3. Results

Our results on predicting emoji sense using our extended Emoji-Lesk algorithm are presented in Table 2. We report our results on both the initial set of senses, as well as the reduced set of senses. We further report our results on a most-frequent-sense baseline, which outperforms our unsupervised results.

In the first row of our results we have shown the initial run using the basic text embedding which just uses the text from the tweet and the basic gloss embeddings. In the following rows, we have then shown the effect on our performance when altering either the text embedding, gloss embedding, or both.

Our results show that some performance gain can be found by incorporating either an embedding for the emoji, or the name of the emoji as text compared to the case where neither is included (the top row of Table 2). This gives extra information during classification and hence leads to the comparative improvement in the results. Interestingly, incorporating hypernyms had little or no positive effect on the classification of our data, however incorporating hyponyms always yielded some increase in classification score.

Our performance is relatively low for the raw data (reported as sim-all in our Table), with a highest similarity of 0.1137 when using the tweet + emoji strategy for the text embedding and the gloss + hyponyms strategy for the gloss embedding. This reflects some small increase over the baseline strategy of just using the tweet and gloss, but not to a significant extent. On the other hand, when using our reduced set of emoji senses, the same set of strategies yields a path similarity score of 0.4146. An average of 0.5 would indicate that we are typically 1 step away in the WordNet hierarchy, so this can be considered a strong and competitive result.

We have also included a Most Frequent Sense baseline, as this is often included in WSD tasks. This is a simplistic baseline in which every item is assigned to the most frequent sense. This strategy performs well in part as it incorporates some extra domain knowledge (the frequency of each sense), which is unavailable to the classification strategy that we have used. The MFS baseline outperforms our systems for both sim-all and sim-reduced. This shows that our systems are approaching the level of a reasonable baseline. Future work may allow us to improve over this baseline as new technologies become available.

Table 3
A sample of errors from our sense prediction task.

Emoji	Tweet	Labelled sense	Predicted sense
Heart eyes	Check out what I just added to my closet…	Beautiful	Message
Heart eyes	DJ HANKKI STOLE MY HEART I ALMOST FORGOT…	Fandom	Quality
Face blowing kiss	I realized Im not getting my email notifications on my work phone!	Apologetic	Quality
Heart	happy birthday angel!!! i hope you are having the best day, i love you	Love	Quality
Heart	Truly lovely to see you recently. Thank you!	Grateful	Message
Heart	Such a great price for such high quality pipes	Pleased	State
Loudly crying	My landlord moved in right next to me. I think he has a keyboard or mini organ.	Annoy	Cognition
Face blowing kiss	first ever blockscreening. thank you everyone for showing your support…	State	Greet
Smiling face with smiling eyes	Who wants to go to Grune tonight and watch some live music with me?	Event	Fashionable

5.4. Error analysis

In Table 3, we have included a few examples of where the sense prediction algorithm did not get the classification correct to help the reader better understand our predictions. These are generated using the best performing predictive system from Table 2 (i.e., tweet + emoji, gloss + hyponyms). In general our predictions were usually close in meaning to the target sense and we have analysed our results to identify both cases where the algorithm had a near miss and those where it got it wrong entirely. In total 25.04% of our predictions were exact matches, and 86.54% of our predictions were within 5 steps of the target label on the WordNet tree.

The examples in Table 3 show that the system usually predicted more generalised labels than those in the labelled senses. This is expected behaviour as the reported senses are those for the sim-reduced category. The most common senses predicted by our system were 'State', 'Quality', 'Message', 'Cognition' and 'Trait' as reflected in the Table, other senses were also predicted for our emoji, but with less frequency. The predicted senses were distributed across the emoji as we expected with our semantic network. This indicates that emoji have senses that transcend their pictographs. In our annotation round we asked our annotators to only return one sense per emoji. However, it may be the case that the emoji bear more than one sense. In this case, our algorithm may have predicted a valid sense, but this is different to the sense returned by the annotator. We may be able to overcome this in the future by allowing annotators to return multiple senses where they feel one sense cannot accurately capture the true meaning of an emoji.

5.5. Emoji sense as a feature

In the previous section, we were able to demonstrate that Emoji can be assigned to a sense index, proving our hypothesis that emoji are polysemous. However, this is only useful when it is used in application. In this section, we investigated the use of our emoji senses as a feature in the emoji prediction task, first proposed at SemEval 2018 (Barbieri et al., 2018). In this task a system is presented with a text and must predict the emoji that was used with this text from a list of N emoji (N = 20 in the original task). The emoji are removed from the texts to prevent the system learning from these. One issue identified in this type of task is that some emoji are indistinguishable from each other. We hypothesise that this is because the emoji share a sense, and are therefore being used synonymously. There is no contextual difference to distinguish between synonymous emoji-senses. For example, consider the 'heart' and 'two hearts' emoji. These are typically used to express love and may be indistinguishable from each other in many contexts. We hypothesise that by including the predicted emoji sense as a feature in the emoji prediction task, we will be able to improve the performance, as the emoji sense will help to distinguish between synonymous emoji categories.

Our study has focused on a different set of emoji than the 20 studied in the shared task in 2018. For this reason, we created our own dataset for predicting 8 emoji using the data described previously in Section 3. To do this, we took all instances of the tweets representing 8 of our emoji (all except 'Black Heart Suit'). We removed the emoji from the texts to prevent contamination of the test set and stored these in a separate file as the numerical index (1–8) of the emoji that each represented as per the original data format of the shared task.

Table 4
The results of incorporating emoji-sense as a feature in the emoji-prediction task.

Features	Accuracy	Macro-F1
Çöltekin et al.	56.5149 ± 0.5391	42.9877 ± 0.4208
Çöltekin et al. + sense	58.8816 ± 0.8517	46.6640 ± 3.2314

We ran the winning system from Semeval 2018 by Çöltekin and Rama (2018) on our data as a baseline and then incorporated the features from our reduced set of synsets (using the best performing combination). The baseline system uses character and word level N-grams as features with an SVM via 5-fold cross validation. We used the default parameters as suggested by the original authors and did not experiment with variations on these to avoid over-tuning to our dataset. We then added the Emoji sense as an extra feature to our SVM representing the sense of the emoji returned by our best predictive model (i.e., tweet + emoji, gloss + hyponyms). This led to a significant improvement ($p < 0.05$) in the macro-F1 on our dataset as shown in Table 4.

6. Discussion

6.1. Why sense-annotate an emoji corpus?

In this paper, we first developed a corpus of tweets each containing one emoji. We selected a sample of tweets and annotated each for the specific sense carried by the emoji. This is the first scientific attempt known to the authors to perform a lexicographic categorisation of the senses borne by emoji. A corpus of this nature can be used for purposes beyond the scope of this research. For example, in sentiment analysis, where emoji and their meanings can drastically affect the sentiment of a text.

By analysing the emoji senses in our corpus, we are able to understand the varied meanings that are carried by each emoji. We are also able to note that emoji are highly polysemous in nature. This is different to words, where one word may have a small collection of senses. Emoji are used without reference to a standardised dictionary, and so the meanings are more varied than words.

6.2. Do the annotators agree?

We found that our annotators disagreed on the annotations in two separate ways. Firstly, the annotators did not always agree on the interpretation of an emoji. This occurred when the context of the emoji was ambiguous, or formed part of a wider dialogue of tweets — which was not present at annotation time. In these cases there may have been one or more legitimate interpretations of an emoji and the annotators rightly chose different senses to reflect these. Secondly, our annotators disagreed on the specific sense key to assign to an emoji, even when the interpretation of that emoji was the same. This is a known issue with the fine-grained sense distinctions in WordNet and we aimed to overcome this through the sense reduction procedure that we have outlined.

Disagreements are present in any annotation task and the general policy is to measure and quantify them with a view to reducing and eliminating disagreement in a final set of annotations. This view relies on the presumption that there is a ground truth that the annotators are aiming for and disagreements arise from lack of training, or understanding the task. However, we take a different view in this work as the assignment of senses is so subjective. Annotators may legitimately disagree on a sense annotation, depending on their own personal cognitive biases. We do not seek to penalise this, but rather to capture and learn from it. This was the reason that we chose to employ a final annotator who we worked closely with to refine the disagreements arising from the crowdsourcing into a final annotation that reflected the research assistant's view of the best possible interpretation, given the senses returned by the crowd workers.

6.3. Can emoji be predicted?

We continued to demonstrate that the emoji senses identified in our corpus could be distinguished from each other by using a relatively simple disambiguation algorithm based on deep learning. This demonstrates that there is some separable information between the classes that we have identified. Our experiments were limited by the small nature of our corpus and further experiments to increase corpus size and perform supervised learning would likely lead to an improvement in predictive score. Notably, our experiments did not surpass a most frequent sense baseline, however this is typically hard to beat in WSD style tasks.

6.4. Are the predicted senses useful for downstream tasks?

We were able to show that in the emoji prediction task, using our sense features helped to classify a text as belonging to one emoji or another. This demonstrates that there is semantic information contained in the text which pertains to certain emoji and not to others. This semantic information helps us to distinguish the sense behind the emoji and therefore select an emoji that bears that sense when predicting the correct emoji for a text.

6.5. How difficult is the overall task?

Determining emoji sense is a difficult task for both machines and humans alike. In our crowdsourcing round, annotators often disagreed on the correct sense and our employed annotator made the final decisions on which senses to assign based on the inputs of the crowd workers. Interpreting the meaning of an emoji requires context. We provided the tweet in which the emoji occurred, however we may also have needed to know the author's intention behind the usage as signified by their common usage of emoji and the discourse in which the emoji occurred. If this task is difficult for a human to do, then it is clearly also difficult for a machine.

In their recent paper, Bender and Koller (2020) argue that learning form and semantics are not the same thing. In our work, we have tried to learn the semantics of emoji, however we can expect that in truth the algorithms that we have used are in fact only identifying features of the form that indicate semantics, rather than semantics themselves. Nevertheless, it is still worthwhile to detect the semantics of emoji as they are widely used, poorly understood and yet very useful for downstream NLP applications.

6.6. How can the task be improved?

As we have stated previously, there is no formalisation of emoji usage. The Unicode Consortium produce emoji and give them names that indicate their intended usage, however these names are not presented in emoji keyboards and people select emoji icons based on their own interpretation of the representation as well as their cultural influences and biases. Different subcultures and geographical regions use the same emoji in different ways and varying emoji dialects may confound applications that are not trained to recognise them. This is an active research problem in socio–linguistics and well beyond the scope of the treatment of emoji in NLP. However, NLP researchers must be aware of this fact when incorporating emoji and tools to recognise their meanings into their pipelines. As with genre transferability in NLP, a system trained on one specific set of usages of emoji is unlikely to be transferable to another set of usages. Further, a system trained to recognise 'general usage' emojis may fall into the trap of overgeneralising and not reflect the specific cultural usages of emoji in a text. This is a hard problem and deserves much future effort from computer scientists and linguists alike.

Our work has focused on a small set of emoji. This is due to limitations in the cost (both financial and time) of annotation of our data. Further work to extend our analysis to more emoji would reinforce the findings we have shown in our paper. Clearly, there is a financial ceiling on doing strict annotations on emoji, so approaches that use automated, or semi-supervised methodologies of annotation may help to speed up this process. Ultimately, we would like to further develop this work to create an 'emoji WordNet' — covering all emoji and the senses that they carry, as well as the links between these senses. Although we have not produced this as part of this current research, the work we have done lays a foundation for future work to attain this goal.

6.7. What is the cost of producing these annotations?

We paid our individual annotators at 3¢ per annotation. In total we gathered 5 judgements for each instance and our corpus contained 200 instances for each of our 9 emoji. This equates to $300 per emoji or $2700 for our entire annotation round. We also employed our research assistant at a rate of 2 days per week (15 h) over 11 weeks at a pay rate of $14.05 per hour. This equates to a further $2318.25. In total our annotations for 9 emoji cost $5018.25.

In our future work we expect to analyse more emoji using a similar methodology, however we will seek to make use of advancement in annotation protocols such as active learning to reduce the overall cost of our emoji annotation problem. Although our annotation costs are high, the benefits of an annotation round are long lasting and the data can be reused in many studies to come.

7. Conclusion

Our work has focused around the central thesis that Emoji are polysemous and their treatment in the NLP literature should reflect this. We have demonstrated this through corpus analysis in Section 4, as well as through empirical analysis in Sections 5 and 5.5. We have proposed a methodology for developing a sense network for emoji, which can be extended to further emoji following the same methodology. This work represents the first piece of research to treat emoji as sememes and we hope that this will inspire future researchers working with emoji to do the same.

CRediT authorship contribution statement

Matthew Shardlow: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Luciano Gerber:** Conceptualization, Methodology, Software, Validation, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Raheel Nawaz:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ahanin, Z., & Ismail, M. A. (2020). Feature extraction based on fuzzy clustering and emoji embeddings for emotion classification. *International Journal of Technology Management and Information System*, *2*, 102–112, URL: http://103.8.145.246/index.php/ijtmis/article/view/8908.

Barbieri, F., Camacho-Collados, J., Ronzano, F., Espinosa-Anke, L., Ballesteros, M., Basile, V., Patti, V., & Saggion, H. (2018). SemEval 2018 task 2: Multilingual emoji prediction. In *Proceedings of the 12th international workshop on semantic evaluation* (pp. 24–33). New Orleans, Louisiana: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/S18-1003, URL: https://www.aclweb.org/anthology/S18-1003.

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proc. of ACL*.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Bond, F., & Foster, R. (2013). Linking and extending an open multilingual WordNet. In *Long papers: Vol. 1*, *Proceedings of the 51st annual meeting of the association for computational linguistics* (pp. 1352–1362).

Budzianowski, P., & Vulic, I. (2019). Hello, it's GPT-2-how can I help you? Towards the use of pretrained language models for task-oriented dialogue systems. In *EMNLP-IJCNLP 2019* (p. 15).

Çöltekin, Ç., & Rama, T. (2018). Tübingen-Oslo at SemEval-2018 task 2: SVMs perform better than RNNs in emoji prediction. In *Proceedings of the 12th international workshop on semantic evaluation* (pp. 34–38). New Orleans, Louisiana: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/S18-1004, URL: https://www.aclweb.org/anthology/S18-1004.

Conneau, A., Kiela, D., Schwenk, H., Barrault, L., & Bordes, A. (2017). Supervised learning of universal sentence representations from natural language inference data. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 670–680). Copenhagen, Denmark: Association for Computational Linguistics, URL: https://www.aclweb.org/anthology/D17-1070.

Delobelle, P., & Berendt, B. (2019). Time to take emoji seriously: They vastly improve casual conversational models. In *Proceedings of the reference AI & ML conference for Belgium, Netherlands & Luxemburg* (pp. 1–7). CEUR Workshop Proceedings.

Devlin, J., Chang, M. -W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies: Vol. 1* (pp. 4171–4186).

Donato, G., & Paggio, P. (2017). Investigating redundancy in emoji use: Study on a Twitter based corpus. In *Proceedings of the 8th workshop on computational approaches to subjectivity, sentiment and social media analysis* (pp. 118–126). Copenhagen, Denmark: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/W17-5216, URL: https://www.aclweb.org/anthology/W17-5216.

Edmonds, P., & Cotton, S. (2001). SENSEVAL-2: Overview. In *Proceedings of SENSEVAL-2 second international workshop on evaluating word sense disambiguation systems* (pp. 1–5).

Eisner, B., Rocktäschel, T., Augenstein, I., Bošnjak, M., & Riedel, S. (2016). Emoji2Vec: Learning emoji representations from their description. In *Proceedings of the fourth international workshop on natural language processing for social media* (pp. 48–54). Austin, TX, USA: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/W16-6208, URL: https://www.aclweb.org/anthology/W16-6208.

Felbo, B., Mislove, A., Søgaard, A., Rahwan, I., & Lehmann, S. (2017). Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 1615–1625). Copenhagen, Denmark: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/D17-1169, URL: https://www.aclweb.org/anthology/D17-1169.

Fellbaum, C. (2010). WordNet. In *Theory and applications of ontology: Computer applications* (pp. 231–243). Springer.

Fellbaum, C., & Vossen, P. (2012). Challenges for a multilingual WordNet. *Language Resources and Evaluation*, *46*, 313–326.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*, 681–694.

Francis, W. N., & Kucera, H. (1979). Brown corpus manual. *Letters to the Editor*, *5*, 7.

Grozea, C. (2004). Finding optimal parameter settings for high performance word sense disambiguation. In *Proceedings of SENSEVAL-3, the third international workshop on the evaluation of systems for the semantic analysis of text* (pp. 125–128). Barcelona, Spain: Association for Computational Linguistics, URL: https://www.aclweb.org/anthology/W04-0831.

Guibon, G., Ochs, M., & Bellot, P. (2018). From emoji usage to categorical emoji prediction. In *19th international conference on computational linguistics and intelligent text processing*. Hanoï, Vietnam: URL: https://hal-amu.archives-ouvertes.fr/hal-01871045.

Huang, L., Sun, C., Qiu, X., & Huang, X. (2019). GlossBERT: BERT for word sense disambiguation with gloss knowledge. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 3509–3514). Hong Kong, China: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/D19-1355, URL: https://www.aclweb.org/anthology/D19-1355.

Hurlburt, G. (2018). Emoji: Lingua Franca or passing fancy? *IT Professional*, *20*, 14–19. http://dx.doi.org/10.1109/MITP.2018.053891332.

Illendula, A., & Yedulla, M. R. (2018). Learning emoji embeddings using emoji co-occurrence network graph. In *Proceedings of the 1st international workshop on emoji understanding and applications in social media*. Stanford, CA.

Kilgarriff, A., & Rosenzweig, J. (2000). Framework and results for English SENSEVAL. *Computers and the Humanities*, *34*, 15–48.

Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on systems documentation* (pp. 24–26).

Ljubešić, N., & Fišer, D. (2016). A global analysis of emoji usage. In *Proceedings of the 10th web as corpus workshop* (pp. 82–89). Berlin: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/W16-2610, URL: https://www.aclweb.org/anthology/W16-2610.

Lu, X., Ai, W., Liu, X., Li, Q., Wang, N., Huang, G., & Mei, Q. (2016). Learning from the ubiquitous language: An empirical analysis of emoji usage of smartphone users. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing* UbiComp '16, (pp. 770–780). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/2971648.2971724, URL: https://doi.org/10.1145/2971648.2971724.

Mihalcea, R., Chklovski, T., & Kilgarriff, A. (2004a). The Senseval-3 English lexical sample task. In *Proceedings of SENSEVAL-3, the third international workshop on the evaluation of systems for the semantic analysis of text* (pp. 25–28).

Mihalcea, R., Chklovski, T., & Kilgarriff, A. (2004b). The senseval-3 english lexical sample task. In *Proceedings of SENSEVAL-3, the third international workshop on the evaluation of systems for the semantic analysis of text* (pp. 25–28). Barcelona, Spain: Association for Computational Linguistics, URL: https://www.aclweb.org/anthology/W04-0807.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, *38*, 39–41.

Na'aman, N., Provenza, H., & Montoya, O. (2017). Varying linguistic purposes of emoji in (Twitter) context. In *Proceedings of ACL 2017, student research workshop* (pp. 136–141). Vancouver, Canada: Association for Computational Linguistics, URL: https://www.aclweb.org/anthology/P17-3022.

Navigli, R. (2009). Word sense disambiguation: A survey. *ACM Computing Surveys*, *41*, 1–69.

Navigli, R., & Ponzetto, S. P. (2010). BabelNet: Building a very large multilingual semantic network. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 216–225).

Niu, C., Li, W., Srihari, R. K., Li, H., & Crist, L. (2004). Context clustering for word sense disambiguation based on modeling pairwise context similarities. In *Proceedings of SENSEVAL-3, the third international workshop on the evaluation of systems for the semantic analysis of text* (pp. 187–190). Barcelona, Spain: Association for Computational Linguistics, URL: https://www.aclweb.org/anthology/W04-0846.

Ramakrishnan, G., Prithviraj, B., & Bhattacharyya, P. (2004). A gloss-centered algorithm for disambiguation. In *Proceedings of SENSEVAL-3, the third international workshop on the evaluation of systems for the semantic analysis of text* (pp. 217–221). Barcelona, Spain: Association for Computational Linguistics, URL: https://www.aclweb.org/anthology/W04-0853.

Sagot, B., & Fišer, D. (2008). Building a free French WordNet from multilingual resources. In *OntoLex*.

Strapparava, C., Gliozzo, A., & Giuliano, C. (2004). Pattern abstraction and term similarity for word sense disambiguation: IRST at Senseval-3. In *Proceedings of SENSEVAL-3, the third international workshop on the evaluation of systems for the semantic analysis of text* (pp. 229–234). Barcelona, Spain: Association for Computational Linguistics, URL: https://www.aclweb.org/anthology/W04-0856.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*, 5998–6008.

Wang, S., & Bond, F. (2013). Building the Chinese open WordNet (cow): Starting from core synsets. In *Proceedings of the 11th workshop on Asian language resources* (pp. 10–18).

Wijeratne, S., Balasuriya, L., Sheth, A., & Doran, D. (2017a). EmojiNet: An open service and API for emoji sense discovery. arXiv:1707.04652.

Wijeratne, S., Balasuriya, L., Sheth, A., & Doran, D. (2017b). A semantics-based measure of emoji similarity. In *Proceedings of the international conference on web intelligence* WI '17, (pp. 646–653). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/3106426.3106490, URL: https://doi.org/10.1145/3106426.3106490.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, *32*, 5753–5763.