Do Emoji Sequences Have a Preferred Word Order?

Susan C. Herring[1] and Jing Ge[2]

Indiana University, Bloomington,[1] University of California, Berkeley[2]
herring@indiana.edu,[1] jingge@berkeley.edu[2]

Abstract

We systematically analyze the ordering of elements in emoji sequences, or uninterrupted strings of two or more different emoji, on Sina Weibo, a prominent Chinese microblogging site. We find that 83% of the sequences can be analyzed as displaying a linear constituent order, with variants of Object-Verb-Subject (OVS) order occurring most frequently. This evidence supports the view that emoji sequences on Sina Weibo have an emergent syntax. Their syntactic properties are not universal, however, since OVS order differs from the Subject-Verb-Object (SVO) order previously reported for emoji sequences in English-language contexts.

Introduction

Emoji usage is increasingly taking on characteristics of verbal language (Herring 2020; Na'aman et al. 2017; Pohl et al. 2017). Some scholars claim that emoji are in fact evolving into a separate, graphical language (Ge and Herring 2018; Monti et al. 2016; Scall 2015). Others further assume that because it is pictographic, emoji language is, or will be, universally understood (Ai et al. 2017).

In this paper, we analyze emoji sequences, or uninterrupted strings of two or more different emoji. Our particular interest is the order in which elements appear in emoji sequences, which can be thought of by analogy with word order patterns in verbal languages (Comrie 1981). If patterned regularities are found in emoji ordering, it would lend strong support to the view that emoji sequences have language-like properties. In particular, it would demonstrate that emoji sequences have syntax, an essential component of language concerned with the arrangement of words, phrases, and clauses in sentences (Moravcsik 2006).

Emoji sequences are posted frequently on Sina Weibo, China's most prominent microblogging site (China Internet Watch 2019), especially by celebrities (Ge and Gretzel 2018). We systematically analyzed the ordering of elements in 300 emoji sequences posted by celebrities and general users on Sina Weibo, asking: To what extent do emoji sequences on Sina Weibo exhibit recurrent patterns in the order of emoji, and which patterns occur most often?

To address these questions, we draw on linguistic research on word order patterns, including the concept of a preferred word order. Preferred word order is the default or pragmatically unmarked order in a language of the core constituents of an independent, declarative clause: subject (S), object (O), and verb (V) (Comrie 1981). We found that more than 80% of the sequences were analyzable as having a linear word order. Verb-subject (VS) and object-verb (OV) orders were most frequent, consistent with a basic OVS order, which together account for 60% of the sequences. The next most frequent overall order, SVO, accounts for only 23% of the sequences. This supports the view that emoji sequences on Sina Weibo have an emergent syntax. Their syntactic properties are not universal, however, since OVS order differs from the SVO order reported for emoji sequences in English-language contexts.

Background

Constituent Order in Emoji Sequences

There is disagreement in the literature as to whether emoji sequences exhibit consistent ordering patterns. Gawne and McCulloch (2019) claim that emoji do not usually combine to form larger structures, and when they do, the positions of the emoji are typically interchangeable, as in 🎄🍰🎁 'Christmas.' In contrast, Schnoebelin, as reported in Steinmetz (2014), asserts that there is a logic to the ordering of emoji in sequences, and that the emoji representing the stance of the speaking subject, which is typically also the grammatical subject, tends to appear in initial position. Danesi (2016) straddles both views, claiming that emoji sequences follow an "iconic-conceptual" syntactic principle while at the same time broadly preserving a basic subject-verb-object order, which he suggests has cognitive, and hence universal, underpinnings. Similarly, in an English-to-emoji translation experiment, Tatman's (2018) subjects used both iconic orders and SVO order. However, Tatman (2018), Danesi (2016), and Schnoebelin's evidence

comes from English-language contexts, and the basic word order of English is subject-verb-object. The possibility cannot be ruled out that emoji users are calquing emoji onto English sentences word-for-word (cf. Danesi, 2016).

The importance of linguistic context becomes apparent when emoji use in non-English contexts is considered. Ge and Herring (2018) reported a tendency in Chinese emoji sequences for the emoji that expresses the speaker's stance to occur in sequence-final position, and also for the emoji that functions like a grammatical object to appear before the verb, suggesting an object-verb-subject order. This is not due to calquing, because the emoji order differs from the basic word order of Chinese, which is subject-verb-object, like English. Stamatov (2017) observed similar tendencies for users of Google Hangouts from the United States and European and Asian countries to create emoji sequences with verb-subject and object-verb orders, e.g., 'Boy rides horse' 🏇👦 (lit. 'ride-horse boy') and 'Look at mobile phone' 📱👀 (lit. 'mobile phone look-at'). Such observations clearly cast doubt on the notion that emoji language is "universal." It is less clear, however, to what extent these patterns are systematic, as opposed to random orders that convey no consistent linguistic meaning.

Word Order Typology

Our working premise is that an emoji sequence is analogous to a sentence-like utterance, functions in some respects like words, and exhibits specific constituent ("word") orders. To explore this premise, we employ a structural linguistics method: word order analysis.

Basic word order refers to the sequence in which the grammatical subject, verb, and object of a transitive sentence occur in oral and written utterances (Russell 1986). Word order typology presents six theoretically possible ways in which the constituents (S, V, O) can be ordered in different languages (Comrie 1981). Basic word order types are summarized and illustrated in Table 1.

Word Orders	Languages	Examples
SVO	English, Mandarin Chinese	They eat pizza.
SOV	Old Chinese, Korean, Turkish	They pizza eat.
VSO	Welsh, Classical Arabic	Eat they pizza.
VOS	Malagasy, Fijian	Eat pizza they.
OVS	Hixkaryana	Pizza eat they.
OSV	Xavante, Warao	Pizza they eat.

Table 1: Word Order Types in Languages (from Comrie 1981).

Method

Data

Emoji sequences are posted frequently on Sina Weibo, especially by celebrities (Ge and Gretzel 2018). As of April

1, 2017, 87 celebrities from a list of top celebrities on Sina Weibo had posted at least one emoji sequence in their most recent 10 posts. Out of all the posts for each of these 87 celebrities, we selected the two most recent stand-alone emoji sequences and the two most recent combinations of sequences and text. In this initial step, 348 emoji sequences were collected. We removed sequences consisting only of repeated emoji, sequences of only emoji with similar meanings (e.g., 😃😄) and sequences consisting of only two emoji in which the second emoji indicates skin tone (e.g., 💃🏻); this left 211 sequences: 102 stand-alone emoji sequences and 109 co-occurrences of emoji sequences and text. We also selected the four most recent user comments that were posted in response to the 211 celebrity posts, of which 20 stand-alone sequences and 41 co-occurrences of sequences and text contained an emoji sequence. Finally, to balance the proportions of stand-alone and dependent sequences, we collected 28 more emoji-only sequences from general users by choosing the first seven sequences among the 211 sequences posted by the celebrities, and then selected the first four user comments involving emoji sequences in response to those seven sequences. Our total dataset thus consists of 300 posts containing emoji sequences that were posted between April 11 and May 28, 2017, of which 150 are stand-alone sequences and 150 are co-occurrences of sequences and text.

The 300 sequences include 1070 individual emoji; 685 are unique occurrences and 385 emoji are repeated. Each sequence consists of an average of 3.6 emoji. We consider both Sina Weibo platform-specific emoji and Unicode emoji (Ge and Herring 2018), because both types are available to Sina Weibo mobile app users, and both are sometimes used together in a single sequence.

Analytical Methods

All 300 emoji sequences in the dataset were manually coded for "word" orders. We first translated the sequences approximately into Chinese words. The second author first translated and coded the entire data set, together with the first author for 70% of the data. Weibo provides definitions of its platform-specific emoji, which proved useful during this process. Following this, 51 ambiguous sequences remained. We then recruited seven coders on Witmart (http://www.witmart.com/cn/), a Chinese crowdsourcing platform similar to Amazon's Mechanical Turk, to provide interpretations of the remaining ambiguous sequences. The crowdsourced coders translated 39 of these sequences with 85% or better interrater agreement. For the remaining 12 unclear examples, we consulted three highly proficient emoji users recommended by a Chinese graphicon design firm. They worked as emoji consultants for the firm and thus were considered expert sources. All 12 ambiguous cases were translated, with 100% agreement among the

coders. Last, the authors generated literal (emoji-by-emoji) and free English translations of all the sequences.

We coded word orders by treating each emoji sequence as a potentially sentence-like utterance and assigning 'subject,' 'verb,' and 'object,' as applicable, to each of the emoji based on its word function in that utterance, taking the discourse context into consideration. Some sequences could not be coded using the word order typology; these were grouped into the categories 'conceptual unit' (CU) for non-clausal sequences or 'other' for cases that did not fit any other category. We considered a non-clausal sequence of at least two semantically different emoji that together express a single concept to be a conceptual unit (e.g., 🍎🥝🍓 'fruit'). The category 'other' was applied to emoji sequences that resemble the "iconic-conceptual" structure noted by Danesi (2016), as well as sequences based on the rebus principle, in which a cluster of emoji represents a word purely based on their pronunciations (Ramsey 1987).

Results

Word Order Frequencies

Counts of "word" orders reveal that most (83.3%) of the emoji sequences were interpretable as having a finite verb plus at least one noun constituent (subject or object) and thus could be classified according to the word order typology in Table 1. (See Table 2.) Transitive clauses with both a surface subject and a surface object comprised 11.3% of the sequences, while in the majority of the clauses either the subject or object was implied, or else the clauses were intransitive and consisted of only a subject and a verb.

The most frequent orders found in our data are OV and VS, which together account for 52.3% of all sequences. Of the sequences that contain all three core constituents, OVS is the most frequent order at 7.3%. In all, roughly 60% of the sequences are consistent with a basic OVS order. SVO (including SV and VO) is the second most frequent order at 22.6%. OV and SV together account for 31.6% of the sequences, although the sequence SOV does not appear.

Category	Number	Percent
OV	81	27%
VS	76	25.3%
VO	45	15%
OVS	22	7.3%
SV	14	4.6%
SVO	9	3%
VOS	3	1%
CU (Conceptual Unit)	44	14.7%
Other	6	2%
Total	300	100%

Table 2: Frequency of Emoji Sequence Word Orders.

Word Order Examples

Examples that illustrate the main patterns in Table 2 are presented and analyzed in what follows.

Example 1 is a stand-alone emoji sequence posted by a celebrity that illustrates repeated emoji functioning as a plural noun and an emoji functioning as a verb:

🎁🎁🎁😋

The first three emoji in the sequence represent 'gifts' (i.e., a plural noun); the last emoji denotes 'love' and functions as a verb. The entire sequence means 'gifts love' (i.e., '(I) love (the) gifts'). Therefore, it displays **OV order**.

The sequence in example 2, posted by a general user, illustrates repeated emoji functioning as an intensified verb and an emoji functioning as a first-person plural pronoun:

🍻🍻🍻🍱

This example was a comment in response to a post by a celebrity regarding his newly released album. The first three emoji in the sequence convey something like 'cheers!!!' and function as an intensified verb (cf. 'to toast someone'); the last emoji indicates the message author and other fans, or 'we' (Ge and Herring 2018). The sequence literally translates as 'toast-toast-toast we' (i.e., 'We heartily toast (you)'). Therefore, it displays **VS order**.

Example 3 is an emoji sequence posted by a celebrity that illustrates two emoji functioning as a verb and a third-person plural form in **VS order**.

乔恩　不　哭, 她们　是　你的　好姐妹,
会　永远　　爱　你. 💕👭
Qiaoen don't cry, they are your sister-like friends,
will forever love you

The Chinese text in example 3 says: 'Don't cry, Qiaoen, they are your sister-like friends and will love you forever.' The first emoji means 'love'; the second emoji indicates the 'sister-like friends' and the 'they' mentioned in the text. The entire sequence means something like 'love they' (i.e., 'They love (you)'). This example is telling, in that the emoji sequence literally translates the main proposition in the preceding text, but whereas the text has normal Chinese SVO order, the emoji version has **VS order**.

Example 4 is an emoji sequence posted by a celebrity that illustrates three emoji functioning as a second-person plural pronoun, a verb, and a first-person singular pronoun:

感谢　亲们　一路　　支持　👪👤💁
Thank dear all all the way support

The text in example 4 translates as: '(I) thank (you,) dear all, for your support all the way'((S)VO). The emoji sequence repeats this message in **OVS order**. The first emoji (defined by Weibo as 'onlookers') here represents the ad-

dressees ('you'); the second emoji denotes 'thank'; and the female gender symbol indicates the speaker ('I'). The sequence means 'you-plural thank I' (i.e., 'I thank you').

Example 5 is a stand-alone emoji sequence posted by a general user that illustrates an emoji functioning as a noun, repeated emoji functioning as a manner verb, and an emoji functioning as a first-person singular pronoun:

This comment was a response to a celebrity post requesting donations of gifts to poor mountain children. The first emoji in the sequence means 'love' and serves as a noun; the string of four emoji means 'deliver/send fast; the last emoji indicates the speaker or message author ('I'). The entire sequence means something like 'love send-fast I' (i.e., 'I send love quickly'). Thus, it displays **OVS order**.

Example 6 is an emoji sequence posted by a celebrity that illustrates two emoji functioning as a verb and a noun:

在 布達佩斯 有 好心的 伯伯 肯 載 偶們
In Budapest have kind   man agree offer ride us

The Chinese text says: 'There was a kind man in Budapest who agreed to offer us a ride.' The first emoji in the sequence means 'see' (i.e., verb), and the second emoji represents a 'van' (i.e., noun). The sequence means 'see van' (i.e., '(We) saw (a) van') and displays **VO order**. The first emoji has a directional orientation (the eyes look to the left); however, it does not follow the emoji that it "sees," making this example somewhat anomalous, especially given the overall preference in the data for OV over VO.

Example 7 is an emoji sequence posted by a celebrity with two emoji functioning as a plural noun and a verb:

和   闺蜜们   的快乐   时光
With girlfriends wonderful time

The text in example 7 means: 'A wonderful time with (my) girlfriends.' The first emoji represents 'girlfriends' as a plural noun; the second emoji denotes the verb 'dance.' The sequence means 'girlfriends dance' (i.e., '(The) girlfriends danced'). Thus it displays **SV order**.

Example 8, posted by a celebrity, illustrates four emoji functioning as a first-person plural pronoun, a verb, a proper noun, and punctuation, respectively, in **SVO order**:

在 我们 心里, 你 就是 胜利者!
In our   heart, you are   the winner!

A translation of the text in example 8 is: 'You are the winner in our heart!' The first emoji in the sequence indicates the speaker and others ('we'); the second emoji means 'love'; the third emoji represents China; and the last emoji serves as a sequence closure marker indicating the speaker's sentiment (Danesi 2016). The entire sequence means 'We love China!' and displays **SVO order**.

Finally, although it was not common in our data, an example of **iconic-conceptual order** is given below. The emoji sequence was posted by a general user and includes emoji that function as pronouns, verbs, and an adjective.

Example 9 was a response to a celebrity's post about her divorce. The first three emoji indicate this female celebrity and her two children; the following two emoji mean 'cheer on'; three 'finger-pointing-to-the-right' emoji fuction as a manner verb (e.g., 'definitely become'); and the last three identical emoji function as an intensified adjective ('happy'). The entire sequence loosely translates as 'You and your children, (we) cheer (you) on, become very happy.'

Discussion

We asked: Do emoji sequences on Sina Weibo display recurrent word order patterns, and if so, what are they? 83.3% of the emoji sequences were analyzable by applying the word order typology, which supports our premise that emoji sequences on Sina Weibo display word order patterns. The orders that occurred most often are OV and SV, consistent with an overall OVS order. Although there is no precise threshold above which a word order can be said to be "basic," the prevalence of OVS order at 60%, more than twice the frequency of any other basic order, makes OVS by far the best candidate for a basic order in our data.

The remaining 16.7% of sequences comprise conceptual units (CUs) in which the order of elements is largely irrelevant, "iconic-conceptual" orders such as example 9, and cases of rebus writing, or the use of symbols to represent words purely for their sounds (Ramsey 1987). All three uses can be considered broadly iconic-conceptual, in that they involve pictographic texts that permit an iconic connection between forms and their referents to be inferred (Danesi 2016). However, even including CU's, this usage accounts for a relatively small portion of our data, in contrast to claims that have been made for emoji sequences in English (e.g., Danesi 2016; Gawne and McCulloch 2019)

That OVS should emerge as the preferred order in emoji sequences on Sina Weibo is a striking result. Not only does it differ from the word order of Chinese, but OVS order is rare, accounting for less than 1% of the world's languages (Gell-Mann and Ruhlen 2011). According to Gell-Mann and Ruhlen, OVS order derives from SOV order, which is the most common word order in the world's languages, and alternates with it, yet there is scant evidence of SOV order in the present data. OV orders are frequent (27%), but they are also consistent with OVS, while SV orders are infrequent (4.6%), and no example of SOV order was found.

When OVS order occurs in verbal languages with a different basic order, it serves marked pragmatic functions

(Gell-Mann and Ruhlen 2011). In contrast, OVS appears to be the unmarked order in Weibo emoji sequences. It is the order used when all the information the sequence contains is old or given from the context, e.g., because the sequence merely repeats the content of the preceding text, as in examples 3 and 4. In these examples, the Chinese text has SVO order, while the emoji "translation" has OVS order. OVS order is also used when all three constituents of the sequence are new or contrasting information, as in example 5. Thus OVS order does not highlight or put focus on any particular constituent in the emoji sequence, a further argument for considering it as the basic or unmarked order.

Ge and Herring (2018) suggest that the placement of emoji representing the subject in sequence-final position is motivated by a Chinese cultural tendency to be self-effacing (cf. Oliver 1969). According to this interpretation, Weibo emoji sequence users foreground actions, activities, and objects and background their own presence, while expressing an attitude or stance that remarks on the sequence as a whole. In some cases, though, the users in our dataset also placed emoji representing first-, second-, and third-person plural pronouns at the end of sequences (examples 2, 3). This suggests that clause-final position is becoming syntacticized as the subject position, independent of cultural motivations. Taken together, the findings from these word order analyses lend support to Ge and Herring's (2018) claim that emoji on Sina Weibo are an "emergent graphical language."

Limitations and Future Directions

We analyzed sequences of emoji used by celebrities and general users in a Chinese language context. Further testing of our findings in other national contexts needs to be done before any general claims about emoji word order can be made. Moreover, in order to better focus on emerging syntactic patterns, we excluded from our sample sequences consisting only of a single repeated emoji. Thus the relative frequencies of sequence types reported here cannot be compared directly with those in previous studies (e.g., McCulloch and Gawne 2018). Finally, although no clear examples were found in our data, some emoji sequences on Weibo appear to contain two verbs, suggesting coordinated or subordinated clause structures. Larger samples of social media discourse that are rich in emoji use should be mined for evidence of these and other incipient syntactic phenomena that were not observable in our limited sample.

References

Ai, W.; Lu, X.; Liu, X.; Wang, N.; Huang, G.; and Mei, Q. 2017. Untangling Emoji Popularity Through Semantic Embeddings. In Proceedings of the Eleventh International AAAI on Web Social Media (pp. 2–11).

China Internet Watch. 2019. Weibo Monthly Active Users Grew Slightly to 497 Million in Q3 2019. https://www.chinainternetwatch.com/30023/weibo-q3-2019-2/

Comrie, B. 1981. *Language Universals and Language Typology. Syntax and Morphology*. Oxford: Blackwell.

Danesi, M. 2016. *The Semiotics of Emoji: The Rise of Visual Language in the Age of the Internet*. London: Bloomsbury.

Ge, J.; and Gretzel, U. 2018. Emoji Rhetoric – A Social Media Influencer Perspective. *Journal of Marketing Management* 34(15-16): 1272-1295.

Ge, J.; and Herring, S. C. 2018. Communicative Functions of Emoji Sequences on Sina Weibo. *First Monday* 23(11). https://firstmonday.org/ojs/index.php/fm/article/view/9413/761

Gawne, L.; and McCulloch, G. 2019. Emoji as Gestures. *Language@Internet* 17, article 2. https://www.languageatinternet.org/articles/2019/

Gell-Mann, M.; and Ruhlen, M. 2011. The Origin and Evolution of Word Order. In Proceedings of the National Academy of Sciences 108(42): 17290–17295. doi: 10.1073/pnas.1113716108.

Herring, S. C. 2020. Grammar and Electronic Language. In *The Concise Encyclopedia of Applied Linguistics*, edited by C. Chapelle. Hoboken, NJ: Wiley-Blackwell.

McCulloch, G.; and Gawne, L. 2018. Emoji Grammar as Beat Gestures. In Proceedings of the 1st International Workshop on Emoji Understanding and Applications in Social Media (Emoji2018). http://ceur-ws.org/Vol-2130/short1.pdf

Monti, J.; Sangati, F.; Chiusaroli, F.; Benjamin, M.; and Mansour 2016. Emojitalianobot and EmojiWorldBot. New Online Tools and Digital Environments for Translation into Emoji. Paper presented at the Third Italian Conference on Computational Linguistics, Naples, Italy, December 5-6.

Moravcsik, E. A. 2006. *An Introduction to Syntax: Fundamentals of Syntactic Analysis*. London/New York: Continuum.

Na'aman, N.; Provenza, H.; and Montoya, O. 2017. Varying Linguistic Purposes of Emoji in (Twitter) Context. In Proceedings of ACL 2017, Student Research Workshop, pp. 136-141.

Oliver, R. T. 1969. The Rhetorical Tradition in China: Confucius and Mencius. *Today's Speech* 17(1), 3-8.

Pohl, H.; Domin, C.; and Rohs, M. 2017. Beyond Just Text: Semantic Emoji Similarity Modeling To Support Expressive Communication. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24(1), article 6.

Ramsey, S. R. 1987. *The Languages of China*. Princeton, NJ: Princeton University Press.

Russell, S. T. 1986. *Basic Word Order: Functional Principles*. London: Croom Helm.

Scall, R. 2015. Emoji as Language and their Place outside American Copyright Law. *New York University Journal of Intellectual Property & Entertainment Law* 5(2), 381-405.

Stamatov, E. G. 2017. Do Emoji Use a Grammar? Emergent Structure in Non-Verbal Digital Communication. Master's thesis, Tilburg University.

Steinmetz, K. 2014. Here are Rules of Using Emoji You Didn't Know You Were Following. *TIME*. http://time.com/2993508/emoji-rules-tweets/

Tatman, R. 2018. Do Emojis Have Their Own Syntax? https://makingnoiseandhearingthings.com/2016/12/07/do-emojis-have-their-own-syntax/