

# Hidden Advertorial Detection on Social Media in Chinese

Meng-Ching Ho<sup>1\*</sup>, Ching-Yun Chuang<sup>2\*</sup>, Yi-Chun Hsu<sup>1\*</sup>, Yu-Yun Chang<sup>1</sup>

<sup>1</sup>Graduate Institute of Linguistics

<sup>2</sup>Master's Program in Digital Content and Technologies (DCT)  
National Chengchi University

109555006@g.nccu.edu.tw, 109462007@g.nccu.edu.tw,

109555003@g.nccu.edu.tw, yuyun@nccu.edu.tw

## Abstract

Nowadays, there are a lot of advertisements hiding as normal posts or experience sharing in social media. There is little research of advertorial detection on Mandarin Chinese texts. This paper thus aimed to focus on hidden advertorial detection of online posts in Taiwan Mandarin Chinese. We inspected seven contextual features based on linguistic theories in discourse level. These features can be further grouped into three schemas under the general advertorial writing structure. We further implemented these features to train a multi-task BERT model to detect advertorials. The results suggested that specific linguistic features would help extract advertorials.

**Keywords:** advertorial, linguistic feature, discourse, advertisement, machine learning

## 1 Introduction

Advertisements on social media are changing, and the vague boundary between advertising and editorial lead to an emergence of a new type of advertising strategy (Keach, 2012). Advertorial is a way of combining advertisement with editorial content. This innovative and hybrid type of discourse involves promotional and journalistic genres, specifically advertisements along with editorials (Deng et al., 2021). Figure 1 is an online



Figure 1: An example of online advertorials posted by an Instagrammer, @imjennim, on date June 3, 2021.

advertorial. This kind of posts often contain advertising cues such as 品牌合作 ‘paid partnership’ and hashtags (e.g., #ad) in the editorials. In addition, utilizing different type sizes and fonts in the layout arrangement are also part of the advertorial tactics (Kim et al., 2001).

According to Keach’s (2012) investigation, people who saw the post through a friend’s recommendation were 24% more likely to view the sponsor positively. Additionally, BuzzFeed, a popular Internet news company, said that advertisements on its site get an average of 40% additional views from people sharing them on social media and that readers are between 10 to 20 times more likely to click on them than on an average banner advertisement (Keach, 2012).

\* These three authors contributed equally.

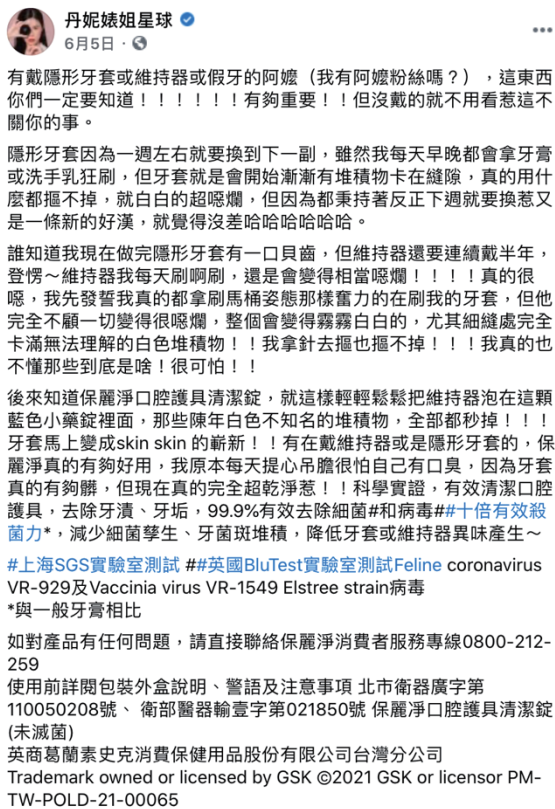


Figure 2: An example of online hidden advertorials posted by a Facebooker, @丹妮婁姐星球, on date June 5, 2021.

Although it is suggested by Guidelines for Editors and Publishers (American Society of Magazine Editors, 2014) that an advertorial should include advertising labels such as “Advertisement”, “Advertising” and “Special Advertising Section”, and the labels should be clearly printed at the page or advertising unit. However, more and more people have ignored and violated the advertising guidelines through time (Ju-Pak et al., 1995). With the lack of these executional cues (e.g., advertisement label, type sizes, and fonts) that act as a way of distinguishing advertorial content from editorial content, we should focus on how advertorials could be identified as such in their contents (Kim et al., 2001). These advertorials that are not clearly labeled are called “hidden advertorials”, and they are contents paid for by advertisers but not identified by viewers. Figure 2 is an example of a hidden advertorial retrieved from Facebook. This advertorial looks like an experience-sharing article, which was hard to be identified at first sight, and does not include clear advertising labels mentioned above.

To the best of our knowledge, little research focused on Chinese advertorial detection,

especially hidden advertorial detection. Due to the overflow of hidden advertorials on social platforms, and the difficulty in distinguishing advertorials from editorials, we aim to observe contextual features from linguistic perspective in discourse level, and apply machine learning models to detect hidden advertorials.

In the study, two research questions are listed in the following: (1) what contextual features are involved in advertorials in Chinese text and their underlying linguistic meanings? and (2) are these contextual features suitable for the classifiers to detect hidden advertorials in Chinese social media?

## 2 Literature review

Hidden advertorials might mislead the public, and the partial and subjective information would influence consumer behaviors. According to a study of Kim et al. (2001), unlabeled advertorials are more likely to be recognized as an editorial material.

Hidden advertising would decrease journalistic quality and cause questioning of the media’s ability to provide the society with reliable and diverse information (Rožukalne, 2010; Rožukalne, 2012). If advertorials are not clearly distinguishable from editorial content in format and tone, legitimate journalism will lose its integrity, and the advertising business could be perceived as a negative image (Ellerbach, 2004). Furthermore, it has been argued that advertorials confuse readers into thinking they are a part of editorial content (Cameron & Haley, 1992). This may indicate that the presence of a label for an advertorial is desired in preventing potential consumer confusion. Therefore, advertising positioned as news must be labeled as “advertorial” to ensure that the readers understand that the message is impartial (Tuten & Perotti, 2019).

To identify advertorials, some relevant studies of features in advertorials are discussed. Persuasive Linguistic Tricks (PLT) (Stepaniuk & Jarosz, 2021) are commonly observed in advertorials. PLT includes strategies such as problems and solutions, which persuades consumers by presenting a solution to a common problem (Stepaniuk & Jarosz, 2021). For instance, hidden advertorials that promote health products and services in magazines usually begin with a life-style problem or a seasonal illness and lead to a problem-solution structure (Kovacic et al., 2012).

Zhou (2012) mentioned advertorials usually employ the strategy into the writing structure, by raising an existing problem in society in the beginning, and then promoting a solution with the technique of product placement to the problem.

Another feature of advertorials are the frequent use of gradable words and expressions (e.g., 非常 ‘extremely’, and 很 ‘very’) that evaluate the products, and take a positive stance toward the sponsors by using positive appreciations (Zhou, 2012). Advertorials usually express positive emotions, hedonic mood and also a great deal of superlative adjectives and nouns to help present a positive image of the sponsors (Zhou, 2012; Burton et al, 2020). Therefore, in advertorials, repetition of positive adjectives and positive verbs are observed. Advertorials also sometimes contain words such as “love” and phrases like “my favorite” to show more assurance from celebrities (Forbes, 2016).

Advertorials also involve some lexical features. The study from Kovacic et al. (2011) combined a textual analysis, in-depth interviews and observations to examine unlabeled advertorials in magazines. The researchers investigated the phenomenon of over-lexicalization in the advertorials, which include several synonymous or near-synonymous terms. Similarly, a linguistic style that measures preciseness is based on whether or not an author uses various terms to describe a single concept (Short & Palmer, 2008). And preciseness would influence the potential to capture viewers (Lee & Theokary, 2021).

In addition, from a large majority of analyzed advertorials, it is observed that advertorials tend to address readers directly. That is, the pronoun “you” is commonly used in advertorials in order to close a ‘discursive gap’ (Fowler, 1991). This constant appearance of “you” is analyzed as the generic use of “you” linguistically, which refers to ‘a specific person the author tries to address to’ or ‘people in general’. The generic use of “you” (henceforth, generic-you) might function as a linguistic nudge that carries persuasive force, which may influence viewers’ behaviors (Orvell et al., 2019).

The study of Labrador et al. (2014) reveals that e-commerce texts typically have two main elements: one to identify the product and the other one to describe it. The latter element includes objective features (e.g., size and weight) and also focuses on persuading the potential customer.

Besides, advertorials often describe the details of the product design, and give advice about their specific use of the product, which allows authors to demonstrate helpfulness and attract more consumers (Forbes, 2016).

Lee’s (2018) investigation based on China market mentioned that hunger-marketing would not only increase customers’ curiosity but also fulfill their conformity. Therefore, one of the PLTs, “uniqueness” (e.g., products or services that are in limited edition), can make the product more desirable by promoting limited editions of the product (Stepaniuk & Jarosz, 2021).

According to a research of Cheung (2010), the discourse strategy, discourse structure and linguistic choices in the sales genre have increased localization. For instance, compared to English advertisements, the results indicated that Chinese advertisements tended to use the strategy, “setting the scene”, which describes the situation and background of the posts, to achieve the social purpose of persuasion more often. Another common persuasive strategy in Chinese advertisements is “offering incentives”. Incentives involved gifts, discounts, free trial or free tests and games. Therefore, aside from the studies of advertorials mentioned above, we intended to focus on advertorials in Taiwan Mandarin Chinese.

### 3 Methodology

This section will be introducing our data collection, feature observation, data preprocessing, and model training.

#### 3.1 Data collection

Our dataset was collected from three online social platforms, which are Dcard, Facebook and Instagram. The dataset consisted of 1,040 articles, and we manually labeled them into 463 advertorials and 577 non-advertorials. The 463 advertorials contained both hidden advertorials and explicit advertorials (with clear advertising labels). Hidden advertorials include posts which are not clearly labeled as advertisements or commercial cooperation, but are authors' own experiences that involve marketing purposes, and can be classified as advertorials based on previous studies and our own observations. The dataset was separated into 70% as a training set (728 posts), 15% as a validation set (156 posts), and 15% as a test set (156 posts). All the articles were in Taiwan



Mandarin Chinese. The reason that we collected articles from three social platforms was that online posts from different platforms may have various advertorial writing styles.

For example, Facebook and Instagram are mostly posted by influencers, and are usually more well-structured. Moreover, sponsors always provide the influencers with special discount codes which can attract more consumers. #AD, #Brand name are common hashtags that label the posts as advertorials on Facebook and Instagram. Dcard usually offers users a free trial of a sponsored product, and users are required to write a review article and share their experience on the platform. Through these review articles, the sponsored products can gain more publicity. Therefore, keywords such as 開箱大使 ‘unboxing ambassador’ and 此文為官方合作文 ‘this post is a collaborative article’ are observed in these review articles which are actually advertorials.

### 3.2 Feature observation

This section introduces the seven features observed in the dataset, which can be further divided into three schemas based on the general advertorial writing structure: background description, product-related information, and building connections with viewers. Features and their related schemas are shown in Table 1.

#### 3.2.1 Background description

Referring to previous studies (Zhou, 2012; Cheung 2010; Kovacic et al., 2012; Stepaniuk & Jarosz, 2021), “setting the scene” is listed as an important feature in our model training. “Setting the scene” includes problem solving and scenario simulation. For instance, skin and hair care problems or social-related issues are commonly seen in collected advertorials.

To identify the posts with the feature “setting the scene”, we extracted posts with keywords such as 問題 ‘problem’, 困擾 ‘problem’ and 煩惱 ‘worries’ and posts that mentioned special events. For instance, 敏感肌問題都跑到臉上作客了嗎? ‘Do you have sensitive skin problems?’ and 即將開學, 穿搭行頭準備好了嗎? ‘Have you prepared your new semester clothing?’

#### 3.2.2 Product-related information

Another significant feature is that these advertorials usually contained very detailed information about the product, which matches the study of Forbes (2016). The detailed information not only includes many technical terms but also elaborated descriptions of the products. Additionally, advertorials related to cosmetics and hair products sometimes includes tutorials and clear steps of how to use the product. This kind of advertorials disguise themselves as instructional content that helps viewers learn how to put on makeup or do hair care routines. There are also some features that we observed from the dataset. Authors often present comparisons of before and after using the product to show the differences and highlight the benefits of the product. In our data processing, we searched for words such as 對比 ‘comparison’ and 前後 ‘before and after’ to filter out the advertorials that implied the effects of the products.

Uniqueness is also a frequent PLT found in our advertorials. In our collected data, keywords like 期間限定 ‘limited time offer’ and 卡友限定 ‘Dcard users only’ which may limit the products to a specific time or group of customers appeared constantly. In our dataset, sponsors often offer free samples and sometimes exclusive discount codes in advertorials, and these incentives have more persuasive potential (Cheung, 2010). Advertisers often raise a ruffle which requires the viewers to leave their comments or tag their friends to boost the reach of the post.

#### 3.2.3 Building connections with viewers

Previous studies showed the influence of positive words in advertorials (Zhou, 2012; Burton, 2020). In the dataset, authors also use exaggeration tones, excessive exclamation marks and various emojis to express their favor and positive appreciation of the products. Two advertorials with this feature are shown in Figure 3 and 4. Furthermore, the direct address of using pronoun ‘you’ may simulate a conversational and relatively personal relationship (Fowler, 1991; Orvell et al, 2019). Additionally, the use of generic-you, which does not refer to a specific person but the general readers, is commonly used in our dataset.

Schemas of writing structure	Features
<i>Background description</i>	(a) setting the scene (including problem solving and scenario simulating)
<i>Product-related information</i>	(b) product-related activities (including ruffles, limited products and free gifts) (c) the comparison of before and after using the product (d) detailed product information (e) product tutorials
<i>Building connections with the viewers</i>	(f) positive and exaggeration tone (g) the use of generic-you

Table 1: The seven contextual features grouped into three schemas under advertorial writing structure for model training.

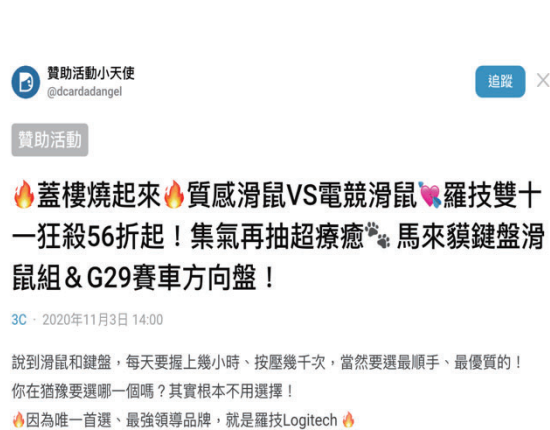


Figure 3: An example of online advertorials using exaggeration tone from Dcard, @dcardangel, on date November 3, 2020.

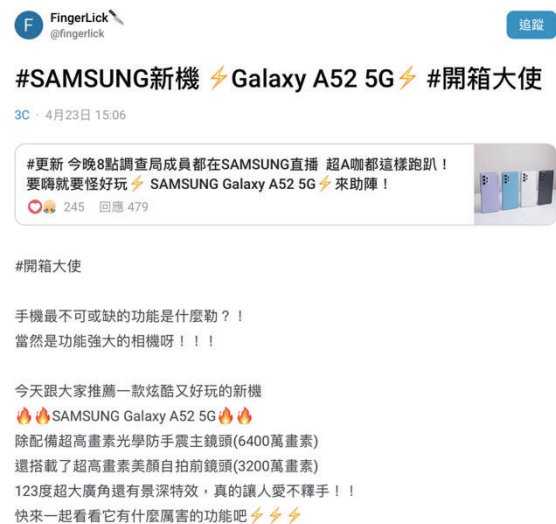


Figure 4: An example of online advertorials using exaggeration tone from Dcard, @fingerlick, on date April 23, 2021.

For example, 說到仙境傳說，你會想到什麼？ ‘What will you think of when it comes to Ragnarok Online Game?’ and IKEA 聽到你的心願了！ ‘IKEA is going to fulfill your dream!’<sup>1</sup>

### 3.3 Data preprocessing and model training

We manually annotated each article with 7 features. We applied one-hot encoding to feature labeling.

We used multi-task learning on Bidirectional Encoder Representations from Transformers (multi-task BERT) referred to Liu et al. (2019) to inspect the effectiveness of the seven features in advertorial detection. Figure 5 shows the architecture of the multi-task BERT model. Multi-task learning is expected to be more beneficial to detect advertorials compared to single task learning. Based on the architecture of multi-task BERT model, we focused on examining the features. We assumed the 7 features as separate classification tasks, and applied multi-task learning on BERT to assess the effectiveness of the features in advertorial detection.

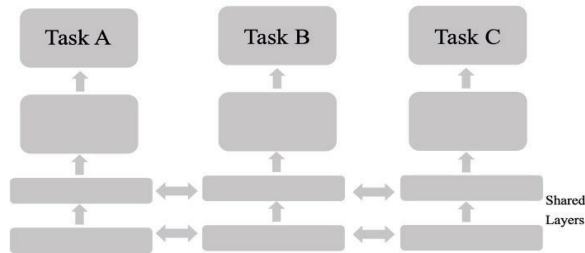


Figure 5: The multi-task BERT model

BERT is a language model based on transformer networks and pre-trained on large corpora. The model uses several multi-head attention layers to learn bidirectional embeddings for input tokens. Some input tokens are masked and the task is to predict a masked word given its context. BERT summed together with positional and segment embeddings using word pieces that passed through an embedding layer (Vaswani et al., 2017; Devlin et al., 2019).

The pre-trained embeddings we used was Chinese-based BERT model (12-layer, 768-

hidden, 12-heads, 103M parameters). The model was trained with 10 epochs, with a learning rate set as  $5e-5$ . The loss weight of a linguist task  $w \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ . After fine-tuning, models have higher performance when  $w$  is in the range of  $[0.3, 0.4]$ . The performance of the model without involving linguistic features was set as a baseline.

## 4 Results

Model performances were evaluated with accuracy, precision, recall, and macro F1 score. The comparison of task performances is shown in Table 2.

We applied all the seven contextual features into the model and the result slightly outperforms the baseline. In order to investigate the effectiveness of each contextual feature to the model performance, we further train the model with seven features respectively.

To compare the model performances of adding various contextual features, different statistical measures (e.g., accuracy, precision, recall, and F1-score) are presented in Table 2. The results indicated that all seven features helped improve the performance of the model. Among all the features, (a) setting the scene, (b) product-related activities, (c) the comparison of before and after using the product, (d) detailed product information, (e) product tutorials, and (g) the use of generic-you appeared to be more effective with F1-score higher than 0.90. Moreover, (g) the use of generic-you had the most significant improvement. On the other hand, (f) positive and exaggeration tone did help advertorial detection but were comparatively less effective than the above mentioned. The results indicated that features would have various degrees of different effectiveness influences on model training.

## 5 Discussion

The results revealed that the features (a) setting the scene, (b) product-related activities, (c) the comparison of before and after using the product, (d) detailed product information, (e) product tutorials, and (g) the use of generic-you may be more effective in advertorial detection. On the

<sup>1</sup> Generic-you are marked in bold type.

Model	Accuracy	Precision	Recall	F1-score
Baseline	0.87	0.87	0.87	0.87
All features	0.88	0.87	0.88	<b>0.88</b>
(a) Setting the scene	0.90	0.90	0.91	<b>0.90</b>
(b) Product-related activities	0.90	0.90	0.89	<b>0.90</b>
(c) The comparison of before and after using the product	0.90	0.91	0.90	<b>0.90</b>
(d) Detailed product information	0.90	0.91	0.90	<b>0.90</b>
(e) Product tutorials	0.90	0.90	0.90	<b>0.90</b>
(f) Positive and exaggeration tone	0.88	0.89	0.87	<b>0.88</b>
(g) The use of generic-you	0.91	0.91	0.91	<b>0.91</b>

Table 2: The comparison of task performances before and after adding contextual features in model training (Boldfaced numbers represented that F1 scores of the features were higher than baseline.)

other hand, (f) positive and exaggeration tone seems to have less effectiveness in identifying advertorials.

Seven features that are categorized into three schemas will be discussed in the following paragraphs. In the schema *background description*, for the feature (a) setting the scene, advertorials usually set up a scene or simulate a scenario in contrast to non-advertorials which express users' experiences only. This may help build connections between the viewers and the authors, which is consistent with previous studies from Cheung (2010), Zhou (2012), and Stepaniuk and Jarosz (2021).

As to the second schema, *product-related information*, which often requires advertorials to provide sufficient information of products, the feature (b) product-related activities can help distinguish advertorials by involving exclusive information and discounts which are seldom seen

in non-advertorials. Another feature (c) the comparison of before and after using the product would be a significant feature to distinguish advertorials. Sponsors especially on Dcard usually recruit users to actually try the products and write review articles. These advertorials often contain the comparison of using the products or not. This kind of comparison can enhance credibility. Other features, (d) detailed product information and (e) product tutorials, can reduce knowledge gaps by effectively communicating this knowledge to the viewers (Burton et al., 2020). Besides, the higher degree of professional knowledge which conveys clear information to their audience can help build credibility and make the information more convincing. The advice about the specific use of the product may aid consumers and allow authors to demonstrate helpfulness and attract more consumers (Forbes, 2016).

Under the third schema, *building connections with the viewers*, the feature (g) the use of generic-you is also commonly seen in advertorials; therefore, it might be a suitable feature to differentiate advertorials from other posts. Generic-you is often used to set a scene or raise a question to build relations with viewers. On the contrary, non-advertorials usually express the authors' own opinions only, which seldom connect to the viewers by using generic-you. This feature is in accordance with previous studies (Fowler, 1991; Kovacic et al., 2011; Orvell et al., 2019). However, the less effective feature (f) positive and exaggeration tone also belongs to the schema *building connections with the viewers*. Positive and exaggerated tone may extract viewers' attention, and the high frequency of emojis and exclamation marks may express the excitement of the authors and thus influence viewers' attitude toward the products. However, people nowadays prefer using lots of emojis to express their emotions even in non-advertorials. Moreover, in the advertorials we collected, to disguise their advertising intentions, the advertorials often avoid using excessive exaggeration tones to be noticed. Therefore, (f) positive and exaggeration tone might be a comparatively less effective feature.

Although in this study, the model trained with all features does not have a higher performance, this might lead to the implementation of the shared layer in the model architecture. Wu et al., (2019) mentioned that since shared features in the shared layer are equally distributed to multi-tasks without filtering, this would introduce noise to the learning process.

We also applied the McNemars' statistical test to compare the performances of the models. It should be noted that the McNemars' test is sensitive to number of counts, and only 156 posts are included in the test set which might bias the result. Although the McNemars' test did not present any significance in our models, the F1-score of the models still indicated that all these independent features may help the model in identifying hidden advertorials.

## 6 Conclusion

This research has attempted to discover the linguistic features in discourse level which can help identify hidden advertorials in Taiwan Mandarin Chinese. We trained a multi-task BERT

model with contextual features based on previous studies and our observations. In general, all the seven contextual features under three schemas mentioned above have enhanced advertorial detection with different degrees of effectiveness.

For future works, larger datasets and other types of genres such as product reviews which are especially difficult to be distinguished from hidden advertorials can be included. Thus, more discourse-level contextual features can be inspected and implemented into machine learning models. Furthermore, different combinations of linguistic features based on schemas can be employed into the models to detect advertorials.

## References

- American Society of Magazine Editors. (2014). ASME Guidelines for Editors and Publishers. In ASME.
- Burton, J. L., Mosteller, J. R., & Hale, K. E. (2020). Using linguistics to inform influencer marketing in services. *Journal of Services Marketing*.
- Cameron, G. T., & Haley, J. E. (1992). Feature advertising: Policies and attitudes in print media. *Journal of Advertising*, 21(3), 47-55.
- Cheung, M. (2010). The globalization and localization of persuasive marketing communication: A cross-linguistic socio-cultural analysis. *Journal of Pragmatics*, 42(2), 354-376.
- Cocker, H., Mardon, R., & Daunt, K. L. (2021). Social media influencers and transgressive celebrity endorsement in consumption community contexts. *European Journal of Marketing*.
- Deng, L., Laghari, T., & Gao, X. (2021). A genre-based exploration of intertextuality and interdiscursivity in advertorial discourse. *English for Specific Purposes (New York, N.Y.)*, 62, 30-42.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Ellerbach, J. (2004). The advertorial as information pollution. *Journal of information ethics*, 61.
- Forbes, K. (2016). Examining the beauty industry's use of social influencers. *Elon Journal of Undergraduate Research in Communications*, 7(2), 78-87.
- Fowler, R. (1991). *Language in the News*. London, New York: Routledge.
- Keach Hagey. (2012). Media: The Advertorial's Best Friend --- BuzzFeed Site Relies on



- Sponsored Content Shared by Visitors on Social Media. *The Wall Street Journal. Eastern Edition*.
- Kim, B. H., Pasadeos, Y., & Barban, A. (2001). On the deceptive effectiveness of labeled and unlabeled advertorial formats. *Mass Communication & Society*, 4(3), 265-281.
- Kovacic, M. P., Erjavec, K., & Stular, K. (2011). Unlabelled advertorials in Slovenian life-style press: A study of the promotion of health products. *Communication & medicine*, 8(2), 157.
- Labrador, B., Ramón, N., Alaiz-Moretón, H., & Sanjurjo-González, H. (2014). Rhetorical structure and persuasive language in the subgenre of online advertisements. *English for Specific Purposes* (New York, N.Y.), 34(1), 38–47.  
<https://doi.org/10.1016/j.esp.2013.10.002>
- Liu, X., He, P., Chen, W., & Gao, J. (2019). Multi-task deep neural networks for natural language understanding. *arXiv preprint arXiv:1901.11504*.
- Orvell, A., Kross, E., & Gelman, S. A. (2019). “You” and “I” in a foreign land: The persuasive force of generic-you. *Journal of Experimental Social Psychology*, 85, 103869–.  
<https://doi.org/10.1016/j.jesp.2019.103869>
- Rožukalne, A. (2012). Significance of hidden advertising of the media business models in Latvia. *Media transformations*, 2012, vol. 8, p. 126-150.
- Ruder, S. (2018, October 24). An Overview of Multi-Task Learning for Deep Learning. Sebastian Ruder. <https://ruder.io/multi-task/>
- Stepaniuk, K., & Jarosz, K. (2021). Persuasive linguistic tricks in social media marketing communication—The memetic approach. *PloS One*, 16(7), e0253983–.
- Tuten, T., & Perotti, V. (2019). Lies, brands and social media. *Qualitative Market Research: An International Journal*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).
- Wu, L., Rao, Y., Jin, H., Nazir, A., & Sun, L. (2019). Different absorption from the same sharing: Sifted multi-task learning for fake news detection. *arXiv preprint arXiv:1909.01720*.
- Zhou, S. (2012). ‘Advertorials’: A genre-based analysis of an emerging hybridized genre. *Discourse & Communication*, 6(3), 323-346.
- 李东昆. (2018). 基于消费心理的饥饿营销策略研究. *中国集体经济*, 2, 66–67.