








MAS: A Corpus of Tweets for Marketing in Spanish

María Navas-Loro¹  , Víctor Rodríguez-Doncel¹ ,
Idafen Santana-Pérez¹ , Alba Fernández-Izquierdo¹ , and Alberto Sánchez²

¹ Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain
{mnavas, vrodriguez, isantana, albafernandez}@fi.upm.es

² Havas Media, Madrid, Spain
alberto.sanchezsf@dbi.io

Abstract. This paper presents a corpus of tweets in Spanish language which were manually tagged for marketing purposes. The used tags describe three aspects of the text of each Twitter post. First, the emotions a brand caused to the author from among a taxonomy of emotions designed by marketing experts. Also, whether it mentioned any element of the *marketing mix* (including various relevant marketing concepts such as *price* or *promotion*). Finally, the position of the author of the tweet with respect to the acquisition process (or *purchase funnel*). Each Twitter post is related to only one brand, which is also indicated in the corpus. The corpus presented in this article is published in a machine-readable format as a collection of RDF documents with links to additional external information. The paper also includes details on the used vocabulary and the tagging criteria, as well as a description of the annotation process followed to tag the tweets.

Keywords: Corpus · Marketing · Marketing mix
Sentiment analysis · NLP · Purchase funnel · Emotion analysis

1 Introduction

Twitter is a source of valuable feedback for companies to probe the public perception of their brands. Whereas sentiment analysis has been extensively applied to social media messages (see [17] among many), other dimensions of brand perception are still of interest and have received less attention [13], specially those related to marketing. In particular, marketing specialists are highly interested in: (a) knowing the position of a tweet author in the purchase funnel (this is, where in the different stages of the customer journey is the author in); (b) knowing to which element or elements of the marketing mix¹ the text refers to and (c) knowing the author's affective situation with respect to a brand in the tweet.

This paper presents the MAS Corpus, a Spanish corpus of tweets of interest for marketing specialists, labeling messages in the three dimensions aforementioned. The corpus is freely available at <http://mascorpus.linkeddata.es/> and has

¹ <http://economictimes.indiatimes.com/definition/marketing-mix>.

been developed in the context of the Spanish research project LPS BIGGER², which analyzed different dimensions of tweets in order to extract relevant information on marketing purposes. A first version of the corpus containing only the sentiment analysis annotations was released as the Corpus for Sentiment Analysis towards Brands (SAB) and was described in [16]. Following this work, we have expanded the corpus tagging the messages in the two remaining dimensions described before: the purchase funnel and the marketing mix. Tweets that were almost identical to others have been removed. Categories of each of the three aspects tagged in the corpus (Sentiment Analysis, Marketing Mix and Purchase Funnel) can be found in Fig. 1.

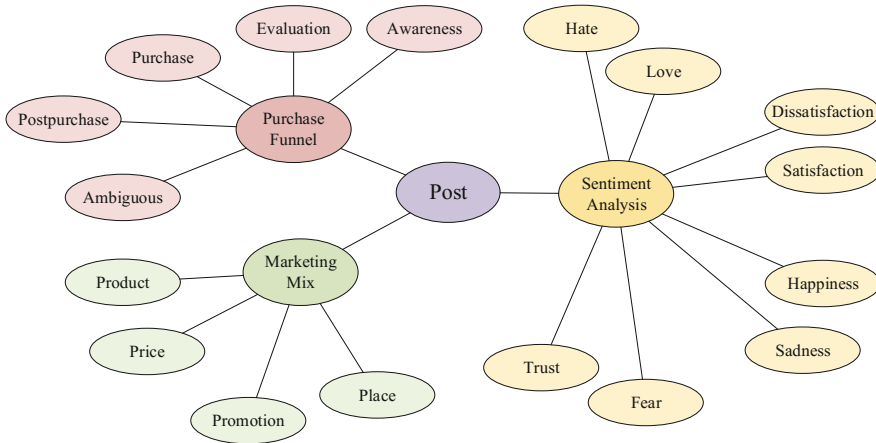


Fig. 1. The three different aspects in an opinion of interest for marketing: Purchase Funnel (*when* is the opinion emitted with regard to the purchase journey), Marketing Mix (*what* is evaluated in the opinion) and Sentiment Analysis (*emotions* expressed with regard to the brand).

2 Related Work

2.1 Sentiment Analysis

Even when Sentiment Analysis is a major field in Natural Language Processing, most of works in Spanish tend to focus on polarity [6,11], being the efforts towards emotions really scarce [23]. Sources of corpora also differ to our aims, since they tend to use specific websites [15,18] and different social networks [6,11,23], and limit to domains such as tourism [15] and medical opinions [18]. Twitter publicly available corpus scarcity is mainly due to two related facts; first, its policy on text dissemination (just the ID of the tweets can be provided, no text nor username). Second, that Twitter periodically deletes tweets

² <http://www.cienlpsbigger.es>.

from their servers; these two facts lead that, even when having the IDs, texts in corpora become impossible to retrieve and therefore the corpus is not available as resource for test anymore. An extended review of works in Spanish Sentiment Analysis with regard to our needs can be found in [16].

2.2 Purchase Funnel

Although different purchase funnel interpretations have been suggested in literature [4, 7], we have based our approach on the one defined in the LPS BIGGER project and already used in [26]. This purchase funnel consists of four different stages (*Awareness*, *Evaluation*, *Purchase* and *Postpurchase*), that reflect how the client gets to know the product, investigates or compares it to other options, acquires it and actually uses and reviews it, respectively.

To the best of our knowledge, there are not public Spanish corpora available containing purchase funnel annotations, since the only work in Spanish on this topic the authors are aware of did not release the dataset used [26]. Nevertheless, the concept of Purchase Intention has been widely covered in literature, especially for marketing purposes in English language. Differently to Sentiment Analysis, Purchase Intention tries to detect or distinguish whether the client intends to buy a product, rather than whether he likes it or not [27]. Starting with the WISH corpus [9], covering wishes in several domains and sources (including product reviews), most works aim to discriminate between different kinds of intentions of users: in [22], the analysis focuses in suggestions and wishes for products and services both in a private dataset and in a part of the previously mentioned WISH corpus; also an analysis performed on tweets about different intentions can be found in [14].

Finally, the most similar categories to the ones in our purchase funnel interpretation are the ones in [5], where the authors differentiate between several kinds of intention, being some of them (such as *wish*, *compare* or *complain*) easy mappable to our purchase funnel stages. Also the corpus used in [10], that classifies into pre-purchase and post-purchase reviews, shares our “timeline” interpretation of the purchase funnel. Out of the marketing domain, corpora labeled with purchase funnel tags for an specific domain have also been published, e.g., for the London musicals and recreational events [8].

2.3 Marketing Mix

Although the original concept of marketing mix [3] contained twelve elements for manufacturers, the most extended categorization for marketing is the one proposed by [12], consisting of four aspects (*price*, *product*, *promotion*, *place*) often known as “the four Ps” (or 4Ps) and revisited several times in literature [25]. Nevertheless, while marketing mix is a well-known and extended concept in the marketing field, in NLP the task of identifying these facets is often simply referred as detecting or recognizing “aspects”, excepting some cases in literature [2]. This task has been often tackled in English [19, 21], while in Spanish corpora we can find a few datasets containing information about aspects, such as those in [6, 20].

3 Tagging Criteria

The corpus consists of more than 3k tweets of brands (see Fig. 1) from different sectors, namely *Food*, *Automotive*, *Banking*, *Beverages*, *Sports*, *Retail* and *Telecom*. When several brands appear in one tweet, just one of them is considered in the tagging process (the marked one); at the same time, the same tweet can appear several times in the corpus considering different brands. Every tweet is tagged in the dimensions exposed in Fig. 1; more than one tag is possible in sentiment and marketing mix dimensions (except simultaneously tagging the pairs of directly opposed emotions), while the purchase funnel, as representing a path on the purchase journey, only presents a tag per tweet. We describe below each dimension, along with a brief report on the criteria used for tagging each category (the complete criteria document, along with statistics on the corpus, can be downloaded with it) (Table 1).

Table 1. Sectors and Brands in the corpus.

Sector	Brand
BEBIDAS (BEVERAGES)	Cruzcampo, Heineken, Estrella Galicia, Mahou
AUTOMOCIÓN (AUTOMOTIVE)	Citroën, Fiat, Hyundai, Kia, Peugeot, Toyota
BANCA (BANKING)	Bankia, Bankinter, BBVA, Sabadell, ING, La Caixa/Caixabank, Santander
ALIMENTACIÓN (FOOD)	Auchan, Bimbo, Hacendado, Milka, Pascual, Puleva
RETAIL	Alcampo, Carrefour, Decathlon, Ikea, Leroy Merlin, Mediamarkt, Mercadona
TELCO (TELECOM)	Amena, Lowi, Movistar, Orange, Vodafone, Yoigo
DEPORTES (SPORTS)	Adidas, Nike, Reebok

3.1 Sentiment Analysis

A tweet can be tagged with one or several emotions, as long as it does not contain directly opposite emotions, or with a *NC2* label meaning there are no emotions on it. Each basic emotion embraces also secondary emotions in it (described in Table 2, and also used by Aguado [1]), and a combination of them can express more complex feelings often seen in customers, such as shown in the following examples:

- When a customer is unable to find a desired product, the post is tagged as *sadness* (for the unavailability) and *satisfaction* (because it reveals previous satisfaction with the brand that deserves the effort of keep looking exactly for it instead of switching to one from another brand).

- When a post shows that a purchase is recurrent, it is tagged as *trust*, referring to the loyalty of the client.
- Emoticons of love are tagged as *love* and musical ones as *happiness* (unless irony happens). *Love* typically implies *happiness*.
- *Happiness* can only be tagged for an already acquired product or service, not for future purchases. The wish for a brand (“*I would like to have X*”) is tagged as *trust* and *satisfaction*, and possibly *love* (depending on the degree of the desire).
- A deep *dissatisfaction* is *hate*, a deep *satisfaction* is *love*.
- *Fear* is understood as the opposite of *trust*.
- The document “*I am satisfied with X but...*” is tagged as *satisfaction*, *trust* plus whichever negative sentiment follows.
- When the emotion is not obvious, the message must be marked as *NC2* (sometimes the meaning is not evident from the message alone). Messages that are evident advertising by the brand community managers, automatic messages or messages related to brand-events are also tagged as *NC2*.
- The emotion must be sparked by the brand/product, and it is not necessarily the same as the message. If the author of a message says that he is sad and he is going to have a *Mahou* (Spanish beer brand), the tag should not be *sadness*, but a positive emotion.
- A deep *dissatisfaction* is *hate*, a deep *satisfaction* is *love*. *Fear* is understood as the opposite of *trust*.

3.2 Purchase Funnel

Each tweet can belong to a stage in the purchase funnel, be ambiguous or be related to a brand without the author being involved in the purchase (such as is the case of posts of the brand itself). Different phases and concrete examples are tagged in the corpus as follows:

Awareness. The first contact of the client with the brand (either showing a willingness to buy or not), usually expressed in first person and mentioning advertising, videos, publicity campaigns, etc. Some examples of awareness would be:

- (1) *I just loved last Movistar ad.*
- (2) *I like the videos in Nike's YouTube channel.*

Evaluation. The post implies some research on the brand (such as questions or seek of confirmation) or comparison to others (by showing preferences among them, for instance), and some interest in acquiring a product or service. Examples of evaluation would be the following:

- (3) *I prefer Citroen to more expensive brands, such as Mercedes or BMW.*
- (4) *Looking for a second-hand Kia Sorento in NY, please send me a DM.*

Table 2. Main emotions and their secondary emotions.

Emotion	Related emotions
Trust	Optimism, Hope, Security
Satisfaction	Fulfillment, Contentment
Happiness	Joy, Gladness, Enjoyment, Delight, Amusement, Joviality, Enthusiasm, Jubilation, Pride, Triumph
Love	Passion, Excitement, Euphoria, Ecstasy
Fear	Nervousness, Alarm, Anxiety, Tenseness, Apprehension, Worry, Shock, Fright, Terror, Panic, Hysteria, Mortification
Dissatisfaction	Dislike, Rejection, Revulsion, Disgust, Irritation, Aggravation, Exasperation, Frustration, Annoyance
Sadness	Depression, Defeat, Hopelessness, Unhappiness, Anguish, Sorrow, Agony, Melancholy, Dejection, Loneliness, Humiliation, Shame, Guilt, Regret, Remorse, Disappointment, Alienation, Isolation, Insecurity
Hate	Rage, Fury, Wrath, Envy, Hostility, Ferocity, Bitterness, Resentment, Spite, Contempt, Vengefulness, Jealousy

Purchase. There is a direct reference to the moment of a purchase or to a clear intention of purchase (usually in first person). Some examples:

- (5) *I've finally decided to switch to Movistar.*
- (6) *Buying my brand new blue Citroen right now!*

Postpurchase. Texts referring to a past purchase or to a current experience, implying to own a product. This class presents a special complexity, since interpretation on the same linguistic patterns change depending on the kind of product, as already exposed in [26] and exemplified in the sentences below:

- (7) *I like Heineken, the taste is so good.*
I would love a Heineken!
- (8) *I like BMWs, they are so classy!*
I would love a BMW!

In (7), the client has likely tasted that beer brand before; people does not tend to like or want beverages they have no experience with (at least without mentioning, such as in “*I want to taste the new Heineken.*”). But the same fact is not derived from more expensive items, even when expressed the same way, such as happens in (8): someone can like a car (such as its appearance or its engine) without having used it or intending to. This is why our criteria states that these kind of expressions must be tagged as *Postpurchase* for some brands (depending on the sector) and others must be tagged as *Ambiguous*, since there can be

several possible and equally likely interpretations. This is kind of expressions has been tagged therefore:

- As *ambiguous* for more expensive products of which people tend to give opinion without acquiring them. In our case, this applies just to the *AUTOMOTIVE* sector.
- As *postpurchase* for products that are easier to acquire, and are likely to have been tried before talking about them. This include the sectors *BEVERAGES*, *BANKING*, *FOOD*, *RETAIL*, *TELECOM* and *SPORTS*.

Ambiguous. This category includes critical posts, suggestions and recommendations, along with posts where it is not clear in which stage the customer is (such as the case mentioned above).

(9) *Do not buy Milka!*

(10) *Loving the new Kia!*

NC2. Includes impersonal messages without opinions (such as corporative news or responses of the brand to clients), questions implying no personal evaluation or intention (for instance, involving a third person), texts with buy or rental offers with no mention to real use experience, etc.

(11) *2008 Hyundai for sale.*

(12) *My aunt didn't like the Kia.*

3.3 Marketing Mix

We have added a *NC2* class to the four original McCarthy's Ps to indicate none of the four aspects is treated in the tweet. It must be noted that, differently than the purchase funnel, several marketing mix tags can appear in the same tweet (except of the *NC2*). Brief explanation of each of the categories tagged for marketing mix, along with examples and part of the criteria, are exposed below:

Product. This category encompasses texts related to the features of the product (such as its quality, performance or taste), along with references to design (such as size, colors or packing) or guaranty, such as in the following examples:

(13) *I find the new iPhone too big for my pocket.*

(14) *I love the new mix Milka Oreo!*

Note that when someone loves/likes something (such as food), we assume it refers to some feature of a product (such as its taste), so we tag it as *Product*.

Promotion. Texts referring to all the promotions and programs of the brand channeled to increase sales and ensure visibility to their products or the brand, such as advertisements, sponsorships (such as prices, sport teams or events), special offers, work offers, promotional articles, etc.

- (15) *Freking out with the new 2x1 @Ikea!*
- (16) *La Liga BBVA is the best league in the world.*

Price. Includes economical aspects of a product, such as references to its value or promotions involving discounts or price drops (that must also be tagged as *Promotion*). Examples of texts that should be tagged as *Price* would be the following:

- (17) *I'm afraid that I can't afford the new Toyota.*
- (18) *Yesterday I saw the same Adidas for just 40e!*

Place. Aspects related to commercialization, such physical places of distribution of the products (for instance, if a product is difficult to find) and customer service (in every stage of the purchase: information, at the point of sale, postpurchase, technical support, etc.).

- (19) *I love the new Milka McFlurry at McDonalds*
- (20) *Already three malls and unable to find the new Nike Pegasus!*

NC2. Impersonal messages of the brand, news or texts that include none of the aspects mentioned before.

- (21) *Nike is paying no tax!*
- (22) *I can't decide between Puleva and Pascual.*

4 The MAS Corpus

4.1 Building the Corpus

A different approach was used for Marketing Mix and the Purchase Funnel tagging with respect to the Sentiment Analysis tagging procedure (where three taggers acted independently with just a common criteria document) exposed in [16]. This meets the need of streamlining the whole tagging process, that happens to be both difficult and time-consuming for taggers. This new procedure is briefly exposed below:

1. A first version of the criteria document was written, based on the study of literature and previous experience within the LPS BIGGER project.
2. Then Tagger 1 tagged a representative part of the corpus (about 800 tweets), highlighting main doubts and dubious tweets with regard to the criteria, that are revised; new tagging examples are added, and some nuances and special cases are rewritten.
3. Taggers 2 and 3 revise the tags by Tagger 1, paying special attention to tweets marked as dubious: if an agreement is reached, the tagging is updated consequently; otherwise, the tweet is tagged as *Ambiguous* or *NC2*.
4. Then each tagger takes a part of the corpus to tag it following the new criteria and highlighting doubts again; these tweets will be revised with remaining taggers, reaching an agreement on the final unique tags in the corpus.

4.2 Publishing the Corpus as Linked Data

We maintain the RDF representation used in the previous version of the corpus, using again our own vocabulary³ to express the purchase funnel and the marketing mix. We also reuse Marl [28] and Onyx [24] for emotions and polarity, and SIOC⁴ and GoodRelations⁵ for post and brand representation. Also links to the entries of brands and companies in external databases such as Thomson Reuters' PermID⁶ and DBpedia⁷ extend the information in the tweets. Figure 2 shows an example of a tweet tagged in the dimensions extracted from the corpus, while Fig. 3 shows information on the company. An overview of this example is depicted in Fig. 4.

4.3 Corpus Description

Final corpus contains 3,763 tweets. Statistics on linguistic information in the corpus can be found in Table 3, along with specific data relevant for Social Media, such as the amount of hashtags, user mentions and URLs. The distribution of categories varies depending on the sector. Mentions of *Place* are for instance more common in *Sports* than in other categories, such as *Beverages* or *Telecom*, as shown in Table 4. Also when opinions are expressed differs: tweets in the *Food*

Table 3. Total and average (per tweet) statistics on the corpus. Stanford CoreNLP was used for POS information, while patterns were used for detecting hashtags ('#'), mentions ('@') and URLs ('*www.**'/'*http**').

	TOTAL	AVG
Tweets	3,763	–
Sentences	5,189	1.38
Tokens	59,555	15.83
Hashtags	1,819	0.48
Mentions	2,306	0.61
URLs	2,111	0.56
Verbs	6,971	1.85
Nouns	8,353	2.22
NPs	6,952	1.85
Adjectives	2,761	0.73
Adverbs	1,584	0.42
Neg. adverbs	560	0.15

³ <http://sabcorpus.linkeddata.es/vocab>.

⁴ <https://www.w3.org/Submission/sioc-spec/>.

⁵ <http://purl.org/goodrelations/>.

⁶ <https://permid.org/>.

⁷ <http://dbpedia.org/>.

Table 4. Statistics on the corpus for the Marketing Mix dimension. Each column represents the percentages of *NC2* (none of the others) and each of the four Ps described in Sect. 3, namely *Product* (PROD), *Price* (PRI), *Promotion* (PROM) and *Place* (PLA)

	NC2	PROD	PRI	PROM	PLA
FOOD	48.80	30.84	2.10	15.27	7.49
AUTOMOTIVE	77.56	4.67	2.00	16.00	1.56
BANKING	53.33	8.50	7.83	21.17	13.17
BEVERAGES	19.85	70.37	2.22	8.59	8.59
SPORTS	54.98	6.43	17.76	0.92	30.32
RETAIL	72.17	12.56	2.09	8.62	7.51
TELECOM	91.63	1.26	1.67	4.60	0.00

Table 5. Statistics on the corpus for the Purchase Funnel categories. Each column represents the percentages of each of the tags described in Sect. 3, namely *NC2* (none of the others) *Evaluation* (EVA), *Awareness* (AWA), *Purchase* (PUR), *Postpurchase* (POS) and *Ambiguous* (AMB).

	NC2	AWA	EVA	PUR	POS	AMB
FOOD	43.41	3.59	3.29	4.19	40.72	5.09
AUTOMOTIVE	85.56	2.67	4.00	0.22	4.44	3.33
BANKING	58.50	5.83	2.00	0.00	7.83	25.67
BEVERAGES	33.63	0.44	13.33	8.44	11.26	32.74
SPORTS	63.09	2.91	4.29	1.84	7.50	19.75
RETAIL	89.29	2.71	4.80	0.62	1.97	1.60
TELECOM	94.14	0.42	0.42	0.00	4.60	0.00

Table 6. Statistics on the emotions in the corpus. Column *ANY* shows the percentage of posts with any emotion (this is, non neutral posts); remaining columns show the percentage of each category among these non neutral posts: *Hate* (HAT), *Sadness* (SAD), *Fear* (FEA), *Dissatisfaction* (DIS), *Satisfaction* (SAT), *Trust* (TRU), *Happiness* (HAP) and *Love* (LOV).

	ANY	HAT	SAD	FEA	DIS	SAT	TRU	HAP	LOV
FOOD	54.79	1.50	1.20	0.00	8.08	45.21	44.01	14.67	12.87
AUTOMOTIVE	9.11	0.00	0.22	1.11	2.44	6.89	3.33	1.11	0.89
BANKING	24.67	5.33	1.00	15.00	23.83	1.33	0.50	0.00	0.00
BEVERAGES	63.11	2.07	1.19	0.74	19.11	44.00	32.74	7.26	7.70
SPORTS	34.15	2.45	2.60	0.31	13.32	18.84	11.94	4.90	11.33
RETAIL	33.00	3.20	1.11	1.48	11.95	14.53	14.41	3.69	3.45
TELECOM	40.17	12.97	0.84	0.00	30.13	8.79	6.28	3.35	1.26

```

mas:827146264517165056 a sioc:Post ;
sioc:id "827146264517165056" ;
sioc:content "Las camisetas nike 2002~2004 y las adidas 2006~2008
son el amor de mi vida"@es ;

marl:describesObject mas:Nike ;
sabd:isInPurchaseFunnel sabv:postPurchase;
sabd:hasMarketingMix sabv:product;
onyx:hasEmotion sabv:love, sabv:satisfaction, sabv:happiness ;
marl:hasPolarity marl:positive ;
marl:forDomain "SPORT" .
    
```

Fig. 2. Sample tagged post “Nike’s 2002–2004 T-shirts and Adidas’ 2006–2008 are the love of my life”, with information such as the tweet ID (*sioc:id*), the text (*sioc:content*, if available), emotion and polarity expressed towards the brand (*marl* and *onyx*), and Purchase Funnel and Marketing Mix tags (*sabd*).

```

mas:Nike a gr:Brand ;
rdfs:seeAlso <http://dbpedia.org/resource/Nike> ;
sabd:1-5000062703 a gr:Business ;
rdfs:label "Nike Inc", "Nike" ;
owl:sameAs permid:1-4295904620 .
    
```

Fig. 3. Extra information on a brand (Nike) and its company (Nike Inc).

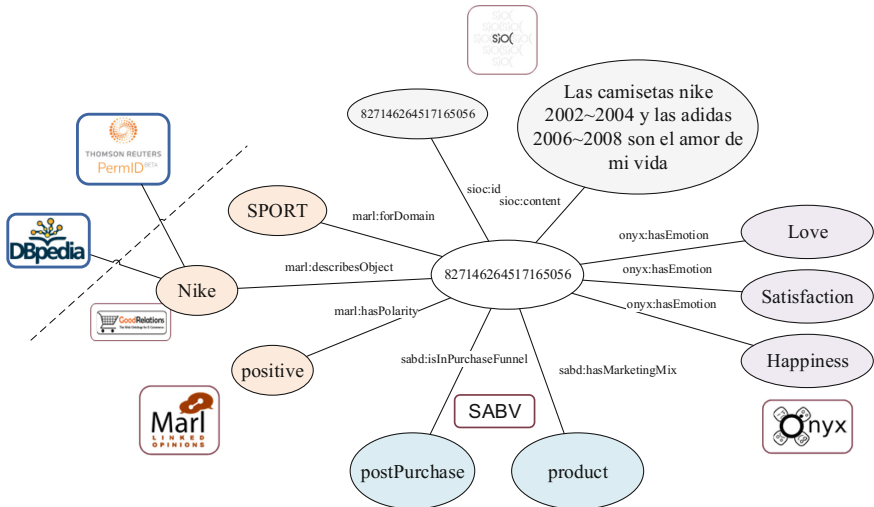


Fig. 4. Example of a tweet using our vocabulary and reusing ontologies (*marl*, *onyx*, *good relations* and *sioc*) and linking to external resources (*DBpedia* and *PermID*) as exposed in Sect. 4.2.

sector tend to refer to the *Postpurchase* phase, while others tend to be more ambiguous or refer to previous phases, as shown in Table 5. Regarding emotions, some of them just appear in certain domains, such as *Fear* for *Banking*, as shown in Table 6.

5 Conclusions

Whereas the SAB corpus provided a collection of tweets tagged with labels useful for making Sentiment Analysis towards brands, this new corpus is of interest for the marketing analysis in a broader way. Using categories specially designed for marketing, such as the four Ps, the MAS Corpus allows marketing professionals to have additional information of habits and behaviors, strong and weak points of the whole purchase experience, and also full insights on concrete aspects of each client reviews.

Acknowledgments. This work has been partially supported by LPS-BIGGER (IDI-20141259), esTextAnalytics project (RTC-2016-4952-7), a Predoctoral grant by the Consejo de Educación, Juventud y Deporte de la Comunidad de Madrid partially founded by the European Social Fund, two Predoctoral grants from the I+D+i program of the Universidad Politécnica de Madrid and a Juan de la Cierva contract. We would also want to thank Pablo Calleja for his help in corpora statistics extraction.

References

1. Aguado, G., et al.: Análisis de sentimientos de un corpus de redes sociales. In: Actas del 31er Congreso Asociación Española de Lingüística Aplicada. Comunicación, Cognición y Cibernética, pp. 522–534 (2013). <http://oa.upm.es/20092/>
2. Bel, N., Diz-pico, J., Pocostales, J.: Classifying short texts for a Social Media monitoring system. Clasificación de textos cortos para un sistema monitor de los Social Media. *Procesamiento del Lenguaje Nat.* **59**, 57–64 (2017)
3. Borden, N.H.: The concept of the marketing mix. *J. Advertising Res.* **4**(2), 2–7 (1964)
4. Bruyn, A.D., Lilien, G.L.: A multi-stage model of word-of-mouth influence through viral marketing. *Int. J. Res. Mark.* **25**(3), 151–163 (2008)
5. Cohan-Sujay, C., Madhulika, Y.: Intention analysis for sales, marketing and customer service. In: Proceedings of COLING 2012, Demonstration Papers, pp. 33–40, December 2012
6. Cumbreiras, M.Á.G., Cámara, E.M., et al.: TASS 2015 - The evolution of the Spanish opinion mining systems. *Procesamiento de Lenguaje Nat.* **56**, 33–40 (2016)
7. Elzinga, D., Mulder, S., Vetvik, O.J., et al.: The consumer decision journey. *McKinsey Q.* **3**, 96–107 (2009)
8. García-Silva, A., Rodríguez-Doncel, V., Corcho, Ó.: Semantic characterization of tweets using topic models: a use case in the entertainment domain. *Int. J. Semantic Web Inf. Syst.* **9**(3), 1–13 (2013)
9. Goldberg, A.B., Fillmore, N., Andrzejewski, D., Xu, Z., Gibson, B., Zhu, X.: May all your wishes come true : a study of wishes and how to recognize them. In: Proceedings of Human Language Technologies: NAACL 2009 (June), pp. 263–271 (2009)

10. Hasan, M., Kotov, A., Mohan, A., Lu, S., Stieg, P.M.: Feedback or research: separating pre-purchase from post-purchase consumer reviews. In: Ferro, N., et al. (eds.) ECIR 2016. LNCS, vol. 9626, pp. 682–688. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30671-1_53
11. Martínez-Cámara, E., Martín-Valdivia, M.T., et al.: Polarity classification for Spanish Tweets using the COST corpus. *J. Inf. Sci.* **41**(3), 263–272 (2015)
12. McCarthy, E.: *Basic Marketing, A Managerial Approach*, 6th edn. Richard D. Irwin, Inc., Homewood (1978)
13. Moghaddam, S.: Beyond sentiment analysis: mining defects and improvements from customer feedback. In: Hanbury, A., Kazai, G., Rauber, A., Fuhr, N. (eds.) ECIR 2015. LNCS, vol. 9022, pp. 400–410. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-16354-3_44
14. Mohamed, H., Mohamed, S.G., Lamjed, B.S.: Customer intentions analysis of twitter based on semantic patterns, pp. 2–6 (2015)
15. Molina-González, M.D., Martínez-Cámara, E., et al.: Cross-domain sentiment analysis using Spanish opinionated words. In: Proceedings of NLDB, pp. 214–219 (2014)
16. Navas-Loro, M., Rodríguez-Doncel, V., Santana-Perez, I., Sánchez, A.: Spanish corpus for sentiment analysis towards brands. In: Proceedings of the 19th International Conference on Speech and Computer (SPECOM), pp. 680–689 (2017)
17. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: LREc, vol. 10 (2010)
18. Plaza-Del-Arco, F.M., Martín-Valdivia, M.T., et al.: COPOS: corpus of patient opinions in Spanish. Application of sentiment analysis techniques. *Procesamiento de Lenguaje Nat.* **57**, 83–90 (2016)
19. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: Semeval-2014 task 4: aspect based sentiment analysis, pp. 27–35, January 2014
20. Pontiki, M., et al.: Semeval-2016 task 5: aspect based sentiment analysis. In: Proceedings of SemEval-2016, pp. 19–30. ACL, San Diego, June 2016
21. Pontiki, M., et al.: SemEval-2016 task 5: aspect based sentiment analysis. In: ProWorkshop SemEval-2016, pp. 19–30. ACL (2016)
22. Ramanand, J., Bhavsar, K., Pedanekar, N.: Wishful thinking: finding suggestions and ‘buy’ wishes from product reviews. In: Proceedings of the NAACL HLT 2010 Workshop (CAAGET 2010) (June), pp. 54–61 (2010)
23. Rangel, F., Rosso, P., Reyes, A.: Emotions and irony per gender in facebook. In: Proceedings of Workshop ES3LOD, LREC-2014, pp. 1–6 (2014)
24. Sánchez Rada, J.F., Torres, M., et al.: A linked data approach to sentiment and emotion analysis of twitter in the financial domain. In: FEOSW (2014)
25. Van Waterschoot, W., Van den Bulte, C.: The 4P classification of the marketing mix revisited. *J. Mark.* **56**, 83–93 (1992)
26. Vázquez, S., Muñoz-García, O., Campanella, I., Poch, M., Fisas, B., Bel, N., Andreu, G.: A classification of user-generated content into consumer decision journey stages. *Neural Networks* **58**(Suppl. C), 68–81 (2014). Special Issue on “Affective Neural Networks and Cognitive Learning Systems for Big Data Analysis”
27. Vineet, G., Devesh, V., Harsh, J., Deepam, K., Shweta, K.: Identifying purchase intent from social posts. In: ICWSM 2014, pp. 180–186 (2014)
28. Westerski, A., Iglesias, C.A., Rico, F.T.: Linked opinions: describing sentiments on the structured web of data. In: Proceedings of the 4th International Workshop Social Data on the Web, vol. 830 (2011)