

From *taggare* to *blessare*: verbal hybrid neologisms in Italian youth slang

Maria Pia DI BUONO (mpdibuono@unior.it)

University of Naples “L’Orientale”, Naples, Italy

Hybrid neologisms, formed directly within the language through various operations of word derivation and formation (Aprile, 2005) and characterised by the use of non-native structures, constitute a diverse system of phenomena and a significant part of contemporary Italian (Dardano, Frenguelli, Perna, 2000), with a growing influence on this linguistic system (Frenguelli, 2005; Giovanardi *et al.*, 2003).

This contribution aims to investigate the phenomenon of verbal hybrid neologisms formed from a non-native morpheme, derived from an English verb, through the analysis of their diffusion in youth slang on social media, as the creation and dissemination of Anglo-Italian neologisms and pseudo-anglicisms can be the result of a *bottom-up Anglicization* (Lubello, 2014: 73).

Due to the fact that documenting the language evolution brought by the introduction of neologisms can be challenging in identifying and classifying new formations, but also in representing their meanings, we propose a methodology to support the discovery and the semantic description of neologisms.

As the first step, using the NLTK Python library, we extract the complete list of single-word English verbs from WordNet and apply a set of manually defined derivative rules to create the hybrid patterns. In Italian, hybrid verb neologisms are usually formed applying the 1st pattern conjugation, represented by the highly productive inflectional morpheme *-are* in the infinitive form. The derivative rules, accounting for the process of morphological blending, include any adjustment rules pertaining to Italian word formation that depend on the type of lexical morpheme used as root. Thus, for instance, monosyllabic verbs ending with a consonant and a vowel are adjusted by an elision of the final vowel and the doubling of the remaining consonant, like *lovvare* from *love* (it is worth stressing that this adjustment for monosyllabic verbs is not consistent, as it might be influenced by an interference with native words, e.g., *zonare* from *zone*).

From this procedure, we obtain a list of hybrid formations which present different statuses according to some dictionaries, e.g., Treccani dictionary¹: i. The formation is already recorded into dictionaries (e.g., *taggare*); ii. The word has been already identified but not included in dictionaries (e.g., *shippare* described in the Treccani Web portal in 2019²); iii. The word is not present in dictionaries and has not been discussed in the Treccani Website (e.g., *blessare* and *lovvare*³). Besides these types, we also identify another group that presents an overlap with native forms, e.g., *zappare* that can be the results of a neologism formation from *zap* or a native form which means *hoe the ground* (for the sake of this work, we do not consider the overlapping formations, as we would need a deeper semantic analysis to distinguish native forms from neologism usages).

¹ <https://www.treccani.it/vocabolario/>

² https://www.treccani.it/magazine/lingua_italiana/domande_e_risposte/lessico/lessico_727.html

³ Some of these are reported in other Web portals, like Slengo an user-made Web portal on Italian neologisms <https://www.slengo.it/define/lovvare>

The list of hybrid verbs are employed in combination with some of the hashtags related to the most popular reality shows and TV programs among young people to extract tweets and create our corpus. Indeed, the practice of commenting on reality shows and TV programs through social networks is widespread, especially among the youth, as evidenced by the daily trending topics published by the social media formerly known as Twitter (now X).

We refine the collected data on the basis of the number of occurrences of each candidate and apply a multidimensional approach to describe the candidate semantics. For the most widespread occurrence, e.g., *shippare*, we use an unsupervised approach, i.e., Latent Dirichlet Allocation - LDA (Blei et al., 2003), to obtain natural clusters for such occurrences. LDA allows for observing the data based on spontaneous aggregations, not decided a priori and it can be applied to detect neologisms and novel word senses (Kim, 2022; Matsumoto et al., 2019; Lau et al., 2012). It represents the underlying information in a text by grouping words in a way that each group is representative of a specific topic. We set the number of topics to 3 and 5 and observe the obtained results, namely clusters of keywords that represent each topic with numerical values indicating the relevance of each word within the topic (for example, for *shippare* the most relevant keywords in Topic 0 are "gregorelli", "romeraci", "giulia", "fan", and so on. This suggests that this topic might concern discussions about shipping different couples and their respective fandoms). Furthermore, we can account for the context usage as the similarity of the topics might indicate that tweets have a very uniform linguistic structure. Finally, being derived from English verbs, our candidates can be evaluated and enriched through the use of English external resources such as VerbNet and WordNet to describe their frame and compare and derive the sentence frame. The obtained results allow for an initial assessment of the diffusion of certain hybridisms and their contexts of use to outline their meaning and semantic fields.

References

- Aprile, M. 2005. *Dalle parole ai dizionari*. Bologna: Il Mulino.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993-1022.
- Dardano, M., Frenguelli, G., & Perna, T. 2000. L'italiano di fronte all'inglese alle soglie del terzo millennio. L'italiano oltre frontiera. In *Atti del V convegno internazionale dell' AISLLI (Leuven, 22—25 aprile 1998)*, 1, 31-55.
- Frenguelli, G. 2005. La composizione con elementi inglesi. In Giovanardi, G. (a cura di) *Lessico e formazione delle parole. Studi offerti a Maurizio Dardano per il suo 70° compleanno*, Firenze: Cesati, 159-176.
- Giovanardi, C., Gualdo, R., & Coco, A. 2003. *Inglese-italiano 1 a 1. Tradurre o non tradurre le parole inglesi*. Lecce: Manni Editori.
- Kim, W. 2022. *Distributional Corpus Analysis of Korean Neologisms Using Artificial Intelligence*. PhD Thesis. University of Georgia.
- Lubello, S. 2014. L'itangliano è ancora lontano? Qualche riflessione sull'influsso dell'inglese. *Id. (a cura di), Lezioni d'italiano. Riflessioni sulla lingua del nuovo millennio*. Bologna: il Mulino, 63-84.

Matsumoto, K., Ren, F., Matsuoka, M., Yoshida, M., & Kita, K. 2019. Slang feature extraction by analysing topic change on social media. *CAAI Transactions on Intelligence Technology*, 4(1), 64-71.

Lau, J. H., Cook, P., McCarthy, D., Newman, D., & Baldwin, T. 2012. Word sense induction for novel sense detection. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 591-601).