Applying Behavioural Profiles to Multimodal Discourse Analysis. A Quantitative Analysis of Self-praise Strategies in French and Russian Social-media
Dylan Glynn, Aliya Gimadieva

To cite this version:

HAL Id: halshs-03807442
https://shs.hal.science/halshs-03807442
Submitted on 9 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Dylan Glynn and Aliya Gimadieva

Applying Behavioural Profiles to Multimodal Discourse Analysis

A Quantitative Analysis of Self-praise Strategies in French and Russian Social-media

Abstract: This study reports the results of a proof-of-principle investigation into the application of the behavioural profile approach to discursive multimodal data. The aim is to demonstrate that a combination of usage-feature analysis and multivariate statistics can be employed to identify discourse patterns integrating both verbal and non-verbal communicative strategies. The case study examines self-praise (boasting) of women (on holidays) on social-media (Instagram) in French and Russian. The assumption is that different causes of pride and/or different languages will result in different strategies of self-praise. The study revealed complex yet coherent patterns across the “type of boast”, the “intensity of the pride” and the “discourse strategy” employed. Moreover, many of the interactions were found to be significant despite the relatively small sample size. Having established that this kind of multimodal discursive phenomena can be adequately operationalized for quantitative treatment, extending the principle to include other social dimensions, such as age and gender, should be non-problematic.

Keywords: critical discourse analysis, boasting, multimodality, multivariate statistics, usage-feature analysis

Acknowledgement: The authors wish to thank the two anonymous reviewers for their suggestions and criticisms. All remaining errors and oversights remain the responsibility of the authors.

1 Introduction: Behavioural Profile Approach and
The behavioural profile approach has proven to be one of the most successful methods for the study of corpus data from a cognitive-functional perspective. Although broadly speaking, this kind of approach has been developed independently in various scientific traditions, in linguistics its origins can be found in early Cognitive Linguistics research, promoted especially by Dirk Geeraerts’ research team at the University of Leuven. Seminal studies such as Dirven et al. (1982); Rudzka-Ostyn (1989); Geeraerts et al. (1994, 1999); Gries (1999, 2003) and Grondelaers (2000) laid the foundations for the contemporary methodological paradigm. Currently, the method has been applied to a wide range of lexical, morpho-syntactic, variationist, and diachronic phenomena as well as more recently, cultural, discursive and even critical discursive phenomena. Despite the wide application of the method, to date it has not been extended to the description of multimodal data and to these ends, this study serves as a methodological proof-of-principle.

That the behavioural profile approach is not widely used in multimodal research is all the more surprising given that Geeraerts et al. (1994), arguably the early work that did the most to popularize the method, employed multimodal data. Although the non-linguistic data (images of various articles of clothing in fashion magazines) were not the object of study per se, they were fundamental to the analysis serving as a tertium comparationis for the annotation. Moreover, Geeraerts et al. (1994) sought to explain semantic structure by including socio-regional (lectal) variation as inherent part of semantically motivated linguistic choices. The current study is directly inspired by the spirit of this early research, not only in its methodology and multimodal data and but also in its semantico-sociolinguistic object of study – the socio-cultural variation of self-praise in social-media.

Self-praise, commonly referred to as boasting or bragging, can be treated as a language function and examined as a discursive object of study. In recent years, the importance of self-praise has arguably increased due to the popularity of various online social media. The stylistic characteristics of these new communicative media are relatively unique and the role of boasting in this genre is central. Instagram posting has become one of the most important genres of social media, both for personal and professional use. Arguably, a grammar of “Insta” posting has emerged, and given that there is an inherently narcissistic dimension to the personal Instagram post as a communicative phenomenon, it represents an excellent opportunity to examine the discursive strategies that have developed to express the discursive function of self-praise. The cliché
highly posed “snapshot” of a semi-clad young woman at an exotic beach destination, captioned with some profound or humorous comment is, in fact, a common choice for Instagrammers. The question is can the behavioural profile approach be employed to accurately describe this kind of inherently multimodal discourse?

2 Method: Usage-feature Analysis and Instagram Posts

2.1 Methodology and linguistic behavioural patterns

The behavioural profile approach, and more specifically multifactorial usage-feature analysis, is a qualitative approach to corpus data that takes the principles of sampling from corpus linguistics and the quantification of variationist sociolinguistics and applies them to semantico-functional language questions. Put simply, the method consists in extracting a relatively large sample of a given linguistic phenomenon and manually tagging each occurrence for various socio-functional characteristics of use. This tagging, or annotation, results in a large multifactorial metadata set – the behavioural profile – which is then examined using multivariate statistics in order to identify complex patterns in use/behaviour of the phenomenon under investigation. Importantly, these patterns can be mathematically modelled in order to calculate the descriptive and predictive accuracy of previous theoretical top-down studies but also bottom-up data driven analyses (q.v. Glynn 2009; Gries and Divjak 2009; Gries 2010; Glynn and Fischer 2010; Glynn and Robinson 2014).

2.2 Data, tokenisation and data extraction

The data examined here belong, in fact, to a larger set that includes English. However, the current results are restricted to French and Russian because it was possible to systematically ascertain the socio-cultural origin of the posts in these languages, where this was not systematically possible for English. The kind of socio-cultural and linguistico-pragmatic effects that the study seeks to identify are exactly the kind of phenomena that will vary depending on the culture to which the speaker belongs, but also the level of language fluency the speaker possesses. Since English is the lingua franca of the Internet, both of
these factors – cultural origin and level of fluency – were impossible control with the English data.

Each Instagram post is associated with an open list of hashtags that permit people to find posts. Russian and French translations of the hashtag “summer 2020” were used to retrieve data. The posts were then manually examined and “tokenized” using the criterion that the image be of a solitary woman with accompanying text and that the post should be by someone indefinable as French or Russian who, in all likelihood, is a native speaker of the language. All other occurrences were removed from the sample. In total, 90 occurrences of both Russian and French were retained (180 occurrences in total). Each post was considered as a single occurrence in its entirety. The possibility of using multiple coders to identify more fine-grained instances of boasting within each post was not adopted (q.v. Fuoli 2012, 2015; Glynn 2014d; Berracheche and Glynn 2016; Berracheche 2020).

Although a sample of 90 occurrences per language is small, especially given the complexity and subtlety of the effects that we are seeking to identify, we believe is reasonable for two reasons. Firstly, if the representativity benchmark is that typical of conversation analysis or traditional discourse analysis, then 180 speaker-unique occurrences, controlled for potentially important random variables, is arguably an excellent first step. If the benchmark is quantifiable representativity, then having a small sample will only make identifying significance more difficult (potential type II error) and although far from ideal, not itself problematic, especially in a proof-of-principle study.

2.3 Usage-feature annotation of text and image

The manual annotation of the data is obviously the most difficult, laborious, and crucial step. In order to develop a set of usage-features for annotation, our starting point was the existing research on self-praise and the social emotions such as pride, Lewandowska-Tomaszczyk (1989); Miller et al. (1992); Wu (2011); Speer (2012) and Dayter (2016) arguably represent the authorities on the discourses of self-praise. With respect to social emotions and pride, research such as Fontaine et al. (2013); Glynn (2014a); Krawczak (2014a, 2014b, 2018); Soares da Silva (2020) and Kumamoto and Glynn (to appear) are representative. Much of this latter research employs the behavioural profile approach and so the primary challenge was extending the analytical apparatus to multimodal data that include images.
2.3.1 Mitigating strategies and caption annotation

The accompanying text, the caption, is an integral part of the self-praise speech event on Instagram. It is immediately visible to the audience and is typically kept short and readable at a glance. In other words, it is integrated into the ‘reading’ of the image as a truly multimodal communicative event. The language analysis per se involves identifying a set of discursive strategies employed by the speaker to encode the self-praise. In the literature, there exists a consensus that self-praise is appraised negatively by speakers (Leech 1983; Brown and Levinson 1987; Dayter 2016) and that much of the discourse associated with this function involves strategies for disguising the self-praise intent of the utterance.

Therefore, the language analysis per se involves identifying a set of discursive strategies employed by the speaker to encode the self-praise. Following Dayter’s (2016) terminology, we will refer to these strategies as mitigating strategies. The actual analysis and tagging is informed by a combination of the above-mentioned literature on self-praise and a bottom up approach based on careful reading of the examples in the data set. Due to practical constraints, explanations of the analysis of each variable will not be given. Nevertheless, examples will be presented that represent the usage-feature identified. The weighted Kappa for each variable, where secondary coding was employed, is given in brackets at the end of the description.

Humour

Several different types of humour were identified and annotated. However, due to the sample size, this level of granularity was not possible to include in the results. The most common humour-based mitigating strategies included false-irony, where a speaker “pretends to be joking” and the use of simple humour references to downplay the narcissistic content of the utterance. Example (1) demonstrates what is understood as false irony and example (2) that of downplay humour ($K=0.89$).

(1) *Qui a dit que les sirènes n’existaient pas*?
‘Who says mermaids don’t exist?’

[Caption accompanying an image of the bikini-clad speaker emerging from the waves at a beautiful beach]

(2) *Et toujours le point levé* 🙄
‘And always/still the clenched fist’ 🙄
Hyperbole

This strategy was particularly interesting and ideally deserves further investigation. It could be argued that underlying the use of hyperbole is some kind of false-irony but this is far from certain. In practice, in our sample, the distinction between humour and hyperbole was clear: hyperbole being employed to express excitement. As a mitigating strategy, it is assumed that the aim was to disguise the self-praise in excitement or happiness at the situation (rather than the speaker in that situation). Recall that the data were restricted to posts about summer and so showing one’s Insta followers an “amazing” holiday destination is a somewhat obvious mitigating strategy. Example (3) is typical of the use of hyperbole ($K = 0.85$).

(3) Невероятной красоты озеро
‘Lake of incredible beauty’

Understatement

The use of litotes and various other explicit downplay devices was an important and relatively straightforward strategy. Put simply, the speaker seeks to downplay the narcissistic content of the boast by understating the photograph. Exotic beaches, beautiful vistas, and amazing skies are described in a matter of fact manner, sometimes to humorous effect, but not necessarily so. Example (4) shows how speakers used understatement to downplay the narcissistic impact of their posts ($K = 86$).

(4) Я и весло 😏🌈
‘Me and a paddle’ 😏🌈

[Caption accompanying an image of a fashion photography pose staring meaningfully off into the distance. She is on a pier extending into a lake that, although pleasant enough, does not appear to be of ‘incredible’ beauty. It is noteworthy that the lake is all but incidental in the image]
We have termed this strategy ‘Philosophy’ for want of a better label. This particular strategy is widespread in the posts and may find its origin in the comic strip captions. Typically, the caption is restricted to bite-sized reflection on ‘life’ and/or a quote from literature or philosophy. Identifying the use of this strategy was largely unproblematic and therefore no secondary annotating was used. Example (5) is typical of the use of this strategy.

(5) “Il faut toute la vie pour apprendre à vivre!”
“It takes the whole of life to learn how to live!”
[Caption accompanying an image of a bikini-clad woman posing naturally in front of what appears to be a beautiful stream for swimming with lush green vegetation in the background]

2.3.2 Self-praise type and image annotation

Instagram is primarily an image-sharing platform where text is incidental. Therefore, the basis of the boast is visual and the annotation of the images was intended to identify ‘types’ of boasts. It follows that the type of boast was operationalized in this study through the feature analysis of the semiotic choices in the photograph taken and chosen to put online. The choice of features for the visual encoding of self-praise was entirely bottom-up and the set of tags that were employed was the result of a process of trial and error. The annotation was performed by a single coder since the various usage-features were all explicit and non-problematic to identify. Nevertheless, in future research, it would be ideal to include secondary coders for all qualitative annotation. Three variables are included in the results presented in the study:

State of dress
The choice to present yourself publicly semi-dressed is a socio-culturally salient semiotic choice. Although degrees of nakedness were included in the original analysis, given the small sample size, this degree of granularity was discarded for the current purposes. The resulting distinction was labelled as “Clothed” and “Semi-Clothed”.

Subject pose

This variable includes both the style of the pose of the subject and the prominence of the subject. The style of the pose was tagged as either “Glamour” and “Informal” depending on whether the subject appeared to pose naturally or less naturally, apparently in imitation of fashion photography. The second dimension distinguished portraits from contextual photographs and was labelled as “Portrait” and “Context” accordingly.

2.3.3 Pride intensity and hashtag annotation

In order to avoid circularity and maintain the independence of our variables, the estimation of the intensity of the experience of the emotion of pride was operationalized through the hashtags employed by the speaker. Unlike the text caption, hashtags are not typically read by the viewer, their role being to lead potential viewers to the post. Therefore, where the speaker will seek to mitigate the self-praise in the caption, the choice of hashtags may “reveal” the true nature of the pride. For this reason, the intensity of the pride was analyzed using the choice of hashtags independently from the post itself. For example, hashtags such as “#sexy” and “#самыекрасивыедевушки” ‘the most beautiful girls’ were annotated as a high intensity of pride. Other hashtags with less obvious narcissistic readings, yet still drawing attention to the poster’s appearance, like “#curlyhair”, “#длинныеволосы” ‘longhair’, “#cheveuxlongs” ‘longhair’ were ranked as relatively intense, while more ideational hashtags, such as “#travelgirl” an “#surfgirl”, were ranked as low intensity. An inherent problem with this operationalization was that almost all postings are to some extent inherently boastful, displaying a certain degree of overt pride.

The degree or intensity of the emotion pride was assessed on a 9-point Likert scale with the intention of reducing that granularity to either a 5-point or 3-point ordinal scale, determined by a Kappa repeatability score. The weighted Kappa estimate at an ordinal granularity of 5 was $K=0.47$ and at a granularity of 3 was $K=0.70$. It is believed that the relatively weak scores for reliability are a result of inadequate training of the second coder. The results of this analysis are, nonetheless, employed since errors in the annotation are most likely to produce less clear results and thus should not artificially increase the evaluation of the method or the accuracy of the description. Nevertheless, this particular shortcoming needs to be underlined. Pride 1 is the label used for the lowest level of intensity and Pride-3 for the highest.
3 Results: Behavioural Patterns of Mitigating Strategies, Pride Intensity, Self-praise Type

3.1 Multidimensional patterns of self-praise

In this section, we will examine patterns that emerge from the associations of the type of mitigating strategy employed relative to the type of self-praise and the intensity of the pride experienced. Instead of summaries of the distribution of the usage-features across these three dimensions in terms of frequencies, we employ multiple correspondence analysis, a technique that calculates the relative frequency of possible combinations of the features and displays those associations in a two-dimensional space. Glynn (2014b) offers a detailed description of the application of this technique to language data and Glynn (2008, 2009, 2014c, 2015, 2016) represent examples of its use applied to the results of manually annotated data. For our current purposes, it is enough to understand that visually presented spatial proximity represents degree of association, where “closer” indicates more highly associated and “further” indicates less association. Note also that the size of the data point represents the contribution to the overall behaviour of the data so that larger data points represent more important contribution to the observed structure. Lastly, distance from the two axes (x and y) is indicative of the degree of distinctiveness in the overall pattern that emerges.

In Figure 1, we find the results of the multiple correspondence analysis where the three dimensions of the study, the hashtag-based intensity of pride, the image-based type of self-praise and the caption-based mitigation strategy, are all combined and represented in a two-dimensional space. Three clusters emerge, each associated with one of the degrees of the intensity of pride. These clusters are, of course, not discrete but tendencies that characterize the discursive strategies of speakers.

In the bottom-right of the plot, the type of self-praise event is characterized by the speaker being fully clothed, regardless of whether she chooses a glamour style pose or a more natural pose and regardless of whether the image is a portrait or a more contextualized shot. These self-praise types are associated with the weakest intensity of pride and are highly associated with the mitigating strategy of “Philosophy” where the caption has little or nothing to do with the image and contains a quote or a philosophical observation.
Fig. 1: Patterns of association between self-praise type, pride intensity and mitigating strategy

Turning to the bottom left, the cluster appears to be most highly associated with self-praise events where the speaker chooses to portray an image of herself semi-clothed and yet typically posed informally, regardless of whether the image is a traditional portrait or a contextual portrait. Interestingly, clothed glamour portraits also belong to this cluster. These self-praise events are associated with a medium intensity of pride and the mitigating strategy of “Humour”, either on its own or in combination with “Philosophy” or “Understatement”. Lastly, and most distinctly, at the very top of the plot, the highest degree of pride intensity is associated with the mitigating strategy of “Understatement”. Lying between this combination and the second level of pride and the use of “Humour” are the two glamour-posed semi-clothed self-praise types. That they lie between two clusters indicates that they are equally associated with both. Similarly, to the right, the mitigating strategies of a combination of “Philosophy” and “Understatement” are found between this highly intense pride cluster and the lowest intensity of pride in the bottom right. It appears that there is a cline
from “Understatement” to “Philosophy” with a combined “Understatement-Philosophy” strategy between the two. Also lying between the two clusters, and therefore equally shared by them, are the strategies of “Hyperbole” and combined “Hyperbole-Philosophy”. Again, the systematicity of the associations, relative to the clusters, is quite striking and suggests that interpretable underlying patterns that represent discursive structures have been identified.

**Fig. 2:** Significant multiway associations and anti-associations in the behavioural patterns

Despite the striking and seemingly coherent nature of these results, multiple correspondence analysis is an exploratory technique and great care must be taken not to extrapolate from its results. Firstly, the analysis presented in Figure 1 represents the distribution of the data to an accuracy of only 56%. What is more, this score refers to only the representation of the data in the sample and offers no information on how generalizable the patterns observed are. In order to have some idea as to the reliability of the results, especially with such a small sample, we run a loglinear analysis on the data. This tests for significance across the multiway associations represented in the correspondence analysis.
However, due to the small sample, the granularity of what can be examined in the loglinear analysis must be greatly reduced. Figure 2 presents the results of this second analysis.

Figure 2 presents a mosaic plot that shows all the potential combinatorial cells across the three dimensions, and using the Chi2 test, calculates the significance of association and disassociation. The colours are based on Person residuals where blue cells represent significant association (more frequent than expected viz. the null-hypothesis) and red cells represent anti-association (less frequent than expected). A detailed description is not possible due to practical limitations, but the results make it clear that several key observations in Figure 1 are indeed significant.

4 Conclusion

The application of the behavioural profile approach to multimodal discourse analysis appears to be successful. Not only were interpretable patterns that characterize discursive strategies identified, but also the preference of one strategy over another for a given context of use was revealed. Moreover, some of the associations that make up those behavioural patterns were shown to be significant, demonstrating that even complex small samples, typical of these kinds of data and research questions are amenable to confirmatory analysis. Importantly, not only was it possible to integrate the multidimensional aspect of the study into the feature analysis and the resulting behavioural profile, but the method appears to be perfectly suited to this very aspect of language complexity. Lastly, although far from perfect, the heuristics used in the operationalization of the intensity and the type of self-praise speech event, produced coherent interpretable and significant results. As proof-of-principle, it seems that the usefulness of applying multifactorial feature analysis and the behavioural approach to discourse structure across different yet integrated modes of language has been demonstrated and deserves wider application to multimodal research.
References


Leech 1983. (Page 5)


