Doing not Being: Concrete Language as a Bridge from Language Technology to Ethnically Inclusive Job Ads

Jetske Adams[a,b], Kyrill Poelmans[b], Iris Hendrickx[c], Martha Larson[a,c]

[a]Institute for Computing and Information Sciences, Radboud University, The Netherlands
[b]Textmetrics, The Netherlands
[c]Centre for Language Studies and Centre for Language and Speech Technology,
Radboud University, The Netherlands

Abstract

This paper makes the case for studying concreteness in language as a bridge that will allow language technology to support the understanding and improvement of ethnic inclusivity in job advertisements. We propose an annotation scheme that guides the assignment of sentences in job ads to classes that reflect concrete actions, i.e., what the employer needs people to *do*, and abstract dispositions, i.e., who the employer expects people to *be*. Using an annotated dataset of Dutch-language job ads, we demonstrate that machine learning technology is effectively able to distinguish these classes.

1 Introduction

Ethnic minorities are disadvantaged in the employment market (Zschirnt and Ruedin, 2016; Andriessen et al., 2012), despite laws that protect them. If people read a job advertisement, and get the sense that the employer will not consider their applications fairly, they will not apply (Verwiebe et al., 2016). This chilling effect can compound already existing employment disadvantages. For this reason, it is important to create welcoming and inclusive job ads.

This paper is motivated by the idea that language technology has potential to help identify job ads that are not inclusive and to suggest changes to make them more welcoming. A conventional machine learning approach would ask human annotators to label a large number of job ads as 'inclusive' and 'not inclusive' and train a classifier. However, ethnic minorities themselves must make the final judgement of the difference between welcoming and unwelcoming ads. Given the burden already borne by these groups, we argue that laborious labeling work should be avoided and a higher-level approach to understanding inclusion in job ads is desirable. In this paper, we aim to build a bridge between language technology and inclusive job ads by investigating basic semantic characteristics of predicates. Specifically, we identify concrete vs. abstract language to be important. In the context of job ads, this distinction translates into the difference between what the employer needs a candidate to *do* on the job and who the employer wants the candidate to *be* in terms of their personal traits.

Our study is inspired by work on stereotypes in job ads by Wille and Derous (2017) who found a difference between *behavioral* statements, e.g., 'You are expected to keep confidential information to yourself', which are concrete and describe the job, and *dispositional* statements that express the same requirement abstractly, e.g., 'You are reliable'. Dispositional statements could be interpreted as a personal judgement that reflects a stereotype that ethnic minorities must frequently face and Wille and Derous (2017) found that they discouraged ethnic minority job applicants from applying. We make the case that language technology that could detect the difference between concrete 'doing' and abstract 'being' would make an important contribution to ethnically inclusive job ads.

Our work makes the following contributions:

- We propose that differences in the concreteness of language use (behavioral vs. dispositional) is a key to using language technology to study inclusivity in job ads.

- We introduce an annotation scheme for labeling sentences in job ads with classes related to behavioral and dispositional language.

- We demonstrate the ability of machine learning approaches to distinguish phrases of different concreteness in job descriptions.

This paper summarizes the most important findings of a larger study of ethnic discrimination in Dutch job advertisements by Adams (2022). We also release an annotated dataset as a resource for the research community.[1]

---

[1] https://github.com/Textmetricslab/Doing-not-Being

2 Background and Related Work

In this section, we provide information on the psychological literature that connects inclusivity with language concreteness and discuss previous work on discrimination detection in job ads.

2.1 Language that Activates Meta-stereotypes

Wille and Derous (2017), mentioned in Sec. 1, carried out field experiments to determine how the requirements listed in job ads, and the way in which they are worded, impact ethnic minorities who are seeking jobs. Their work is informed by the concept of a *meta-stereotype*, which was introduced by Vorauer et al. (2000) to describe a trait whose mention triggers a discriminated group to assume they are being stereotyped. The words 'integrity', 'trustworthy', and 'reliable' are given as examples. A study by Bhargava and Theunissen (2019) further demonstrates that ethnic minorities are likely to disassociate with dispositional phrases in job ads. Occupational stereotypes reflected in this wording hinder encouragement of a diverse group of applicants. Wille and Derous (2017) recommend to focus on people's potential to do the job and not on innate traits in the recruitment process. Their work is guided by the Linguistic Category Model (LCM) (Semin and Fiedler, 1991), which organizes verbs and adjectives along a linear scale with verbs (related to behavior) on the concrete side and adjectives (related to disposition) on the abstract side. In our work, the LCM informs the development of our annotation guidelines.

2.2 Language that Creates Distance

Construal Level Theory (Trope and Liberman, 2010) holds that increased psychological distance corresponds to increased abstraction. Detailed, concrete, and descriptive language is associated with small social distance. Abstract language that reflects innate and lasting qualities is associated with large social distance. In a job ad, the same requirement can be formulated with increasing levels of abstraction, suggesting increasing social distance:

1. You advise customers about the use of our products.
2. You are focused on sensing customer needs.
3. You are customer-oriented.

If using formulations that decrease social distance makes a job ad more welcoming, then CLT supports our idea that studying language concreteness can contribute to ethnic inclusivity.

Work that associates high levels of social power with the use of abstract language (Wakslak et al., 2014) provides further support. Assuming that large perceived power distance could be unwelcoming, this work also points towards language concreteness being important for ethnically inclusive job advertisements.

2.3 Technology for inclusive job ads

Work on language technology for studying discrimination in job ads is surprisingly limited. The closest work to our own is Ningrum et al. (2020). This work uses a Discriminatory Keyword Dictionary (DKD) and Word Pattern Templates (WPTs) to detect different types of discrimination in Indonesian job ads. Although this study did not look specifically at ethnic minorities, it did find that direct discrimination on the basis of religion, often correlated with ethnicity, was present in about 1 of 100 job ads. In contrast, we are not interested in detecting discrimination, but instead in detecting phrasing that might trigger job applicants to be concerned that discrimination might be forthcoming. To our knowledge, we are the first to propose to understand and improve the ethnic inclusivity of job ads by way of language technology capable of detecting language concreteness.

3 Method

We first discuss the annotation scheme that converts the class scheme of the LCM to the job advertisement domain and how we applied this to manually label a sample of job advertisement phrases. Then, we describe a supervised machine learning approach on a small set of job advertisement phrases in order to demonstrate that the distinction between dispositional and behavioral phrasing can be automatically detected consistently and accurately as a proof of concept of the applicability of language concreteness estimation in job ads.

3.1 Annotation scheme

We used the LCM to operationalize Construal Level Theory since it offers an implementation of a scale (i.e., continuum) of phrasal expressions from concrete to abstract. Each of the classes proposed by the LCM was adapted to the domain of job ads, both in name and definition. The annotation scheme is summarized in Figure 1. The definitions
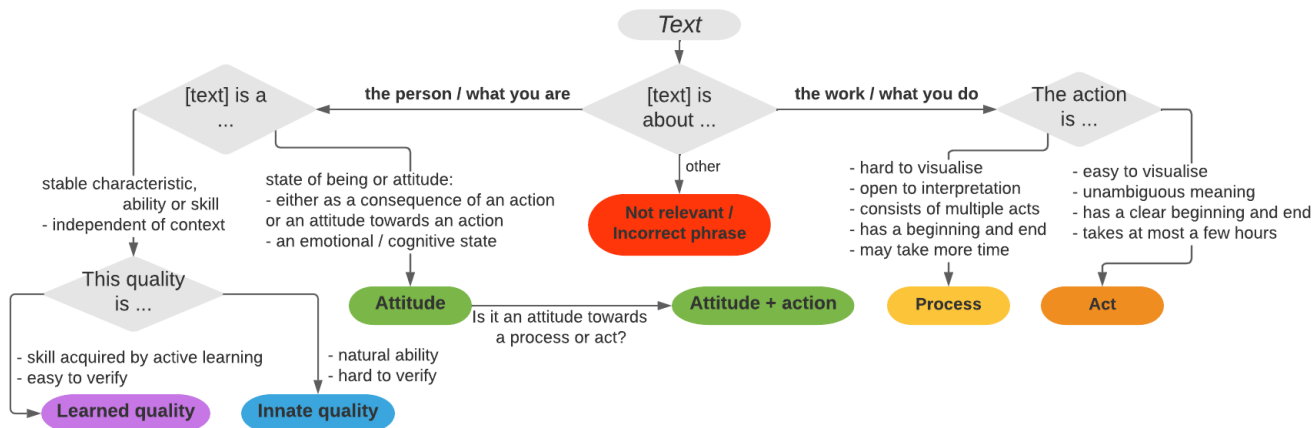
Figure 1: Annotation scheme for Behavioral/Dispositional classes reflecting concrete/abstract language in job ads

of the labels are provided, and elaborated on, in Appendix A. Six sub-classes were defined, with, from most concrete to most abstract, 'Act' and 'Process' as *behavioral classes* and 'Attitude + action', 'Attitude', and 'Innate quality' as *dispositional* classes. 'Learned quality' is added for completeness and taken to be *dispositional*, but not abstract.

3.2 Data and Annotation

Job advertisements contain a very typical language use and structure. As we are interested in advertisements on the Dutch job market, we needed to create a data sample of Dutch job advertisements and develop a set of annotation guidelines to apply the LCM model to our sample. We focus on the annotation of verb predicates and high-frequency domain-specific nouns (such as 'experience' or 'technical aptitude') as these are most likely to describe job requirements and qualities.

We used a sample of 17,810 Dutch advertisements collected in 2021 from diverse job ad platforms and representing different job branches. From this collection, 4,000 sentences were randomly extracted from the middle of the advertisements, where we expected to find mention of job requirements, and were manually annotated according to our annotation scheme (Fig. 1 and 2). The sentences were automatically parsed with Frog, a Dutch NLP tool (van den Bosch et al., 2007).

We are interested in annotating the part of the sentence that constitutes the predicate. To this end we extracted verb phrases and relevant nouns, using a set of rules based on PoS tags, phrase chunks, and dependency relations.

The application of the LCM to job advertisement texts was by no means a trivial task and required an extensive development phase. Development consisted of a series of pilots performed with a group of annotators on a separate sample consisting of job ads collected in 2014. We needed five rounds of annotation pilots to converge to a final version of the annotation guidelines that could be applied with sufficiently high inter-annotator agreement. In total, in our final dataset, 5,277 predicates and nouns were manually annotated by three annotators (Krippendorff's alpha (α) = 0.77).

Figure 2: Examples of manually annotated sentences from the validation set (some were shortened), translated from Dutch and visualized with displaCy[2].

The annotated dataset was split sentence-wise using a stratified random sampling strategy such that the predicates are proportionally balanced over the sub-classes. The data was split with a ratio of 70:15:15, resulting in a training, validation, and test set of respectively 3,654, 788, and 785 predicates. We measure performance on the validation and test set using Area Under the ROC curve (AUROC).

---

[2]https://spacy.io/usage/visualizers

Model	Relevant vs. Not relevant		Dispositional vs. Behavioral		Sub-classes	
	Val	Test	Val	Test	Val	Test
TF-IDF + Naïve Bayes	.85	.87	.93	.92	.90	.90
Word2Vec + LSTM	.89	.89	.94	.93	.92	.91
BERT fine-tuned	.93	.92	.96	.96	.92	.92
RoBERTa fine-tuned	.93	.94	.95	.94	.90	.91

Table 1: Proof-of-concept results on our validation and test sets (Micro-average AUROC scores)

4 Dispositional/Behaviorial Detection

We took a three-step approach to automatically detecting concreteness/abstractness classes. First, the predicates were extracted from the sentences using the rule-based method described in Sec. 3.2. Second, the extracted predicates were classified by their relevance, and discarded if they were not dispositional or behavioral. Third, the relevant predicates were classified by a binary classifier as Dispositional/Behavioral (left/right of Fig. 1) and by a multi-class classifier into the sub-classes (bottom classes of Fig. 1). We evaluated four classifiers:

TF-IDF + Naïve Bayes TF-IDF weighed feature vectors were extracted from the data and dimensionality reduction was applied. (We used variance thresholding at threshold = 0.0005 and the Chi-Square test to reduce the vector size to 500.) Naïve Bayes was implemented using scikit-learn (Pedregosa et al., 2011).

Word2Vec + LSTM pre-trained Dutch word embeddings (320-dimensional) from Tulkens et al. (2016) were used as input to an LSTM, using Python an Keras Tensorflow.

BERT 'BERTje' (de Vries et al., 2019), a Dutch, pre-trained, transformer-based BERT model, was fine-tuned using Python and Keras Tensorflow. A dropout layer was added for regularization.

RoBERTa 'RobBERT' (Delobelle et al., 2020) was fine-tuned in similar fashion.

We also experimented with a (one-step) token classification approach, similar to NER. This resulted in incorrect and spurious predicate detection and was not explored further here.

5 Results

Tab. 1 presents results that confirm the ability of a machine learning approach to distinguish dispositional and behavioral predicates. The neural models (Word2Vec + LSTM, BERT and RoBERTa) outperform Naïve Bayes and the transformer-based models give the best over-all performance.

Fig. 3 presents the confusion matrix of the sub-classes, which yields the following insights:

Error severity Recall that the sub-classes (except 'Learned quality') are placed along a continuum from concrete to abstract. Fig. 3 shows that the incorrectly predicted labels are often close to the ground truth label on this continuum.

Figure 3: Confusion matrix for BERT over the sub-classes (test set predicates in the class 'Relevant').

Class confusion 'Process' is confused most often with 'Act'. During the annotation pilots, it was already observed that it is hard to judge the edge cases between these classes. For example, take the predicate *taking care of the project documentation*. It is not clear-cut to which class this example belongs. The class 'Attitude' is confused with 'Attitude + action' or 'Innate quality'. Phrases of these types are often syntactically similar. The class 'Learned quality' shows the least confusion. This observation is not surprising because 'Learned quality' is the majority class in the data and is most easily identifiable by specific frequently occuring nouns (e.g., names of certificates, education levels, language skills, or words like *ervaring* English: 'experience' or *kennis* English: 'knowledge').

6 Conclusion and Outlook

In this paper, we have proposed that language concreteness is useful as a bridge between language technology and ethnic inclusivity in job ads. The connection between inclusivity and concrete language is supported by research that has shown that focusing on *doing* rather than *being* can prevent ethnic minorities from being put off by job ads that they are qualified to apply for. It is also supported by the psychology literature on social distance and social power distance. We presented an annotation scheme that supports stable annotation of classes along a continuum that runs from abstract (dispositional) to concrete (behavioral) and have used it to annotate a dataset of Dutch-language job ads. The dataset has allowed us to demonstrate that machine learning classifiers can reliably detect differences in language concreteness. We intend our work to be useful to machine learning researchers, who can apply our annotation scheme and reproduce our experiments for different datasets and languages, but especially to social psychologists, as they continue to investigate ethnic inclusivity in the employment market.

It is important to note the difference between our work and other work that has been carried out on ethnic bias in NLP models, e.g., Ahn and Oh (2021) and Nadeem et al. (2021). The concern of these studies is stereotypes that are expressed about members of ethnic minorities. In other words, they focus on the context in which ethnic minorities are mentioned and/or what is said about them. In contrast, our work studies textual phrasing that could trigger members of ethnic minorities to be concerned that the writer may hold stereotypes against them. This contrast is important because whether or not a job ad is perceived as inclusive goes far beyond direct mentions of ethnic minorities. We hope that our work is useful to extend the understanding of how ethnic inclusivity can be promoted in society, and how NLP can contribute to this goal.

7 Acknowledgements

References

Jetske Adams. 2022. Automating the detection of dispositional and behavioural phrasing, the linguistic category model applied to Dutch job advertisements. Master's thesis, Radboud University.

Jaimeen Ahn and Alice Oh. 2021. Mitigating language-dependent ethnic bias in BERT. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 533–549. Association for Computational Linguistics.

Iris Andriessen, Eline Nievers, Jaco Dagevos, and Laila Faulk. 2012. Ethnic discrimination in the Dutch labor market: Its relationship with job characteristics and multiple group membership. *Work and Occupations*, 39(3):237–269.

Deepti Bhargava and Petra Theunissen. 2019. The future of PR is 'fantastic', 'friendly' and 'funny': Occupational stereotypes and symbolic capital in entry-level job advertisements. *Public Relations Review*, 45(4):101822.

Wietse de Vries, Andreas van Cranenburgh, Arianna Bisazza, Tommaso Caselli, Gertjan van Noord, and Malvina Nissim. 2019. BERTje: A Dutch BERT Model. arXiv:1912.09582.

Pieter Delobelle, Thomas Winters, and Bettina Berendt. 2020. RobBERT: a Dutch RoBERTa-based Language Model. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 3255–3265, Online. Association for Computational Linguistics.

Moin Nadeem, Anna Bethke, and Siva Reddy. 2021. StereoSet: Measuring stereotypical bias in pretrained language models. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5356–5371, Online. Association for Computational Linguistics.

Panggih Kusuma Ningrum, Tatdow Pansombut, and Attachai Ueranantasun. 2020. Text mining of online job advertisements to identify direct discrimination during job hunting process: A case study in Indonesia. *PLOS ONE*, 15(6):e0233746.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Gün R. Semin and Klaus Fiedler. 1991. The Linguistic Category Model, its bases, applications and range. *European Review of Social Psychology*, 2(1):1–30.

Yaacov Trope and Nira Liberman. 2010. Construal-level theory of psychological distance. *Psychological Review*, 117(2):440.

Stéphan Tulkens, Chris Emmery, and Walter Daelemans. 2016. Evaluating unsupervised Dutch word embeddings as a linguistic resource. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Antal van den Bosch, Bertjan Busser, Sander Canisius, and Walter Daelemans. 2007. An efficient memory-based morphosyntactic tagger and parser for Dutch. *Computational Linguistics in the Netherlands: Selected Papers from the Seventeenth CLIN Meeting*, pages 99–114.

Roland Verwiebe, Lena Seewann, Margarita Wolf, and Melek Hacioglu. 2016. 'I have to be very good in what I do'. Marginalisation and discrimination in the career-entry phase–experiences and coping strategies among university graduates with a migrant background in Austria. *Journal of Ethnic and Migration Studies*, 42(15):2468–2490.

Jacquie D. Vorauer, A. J. Hunter, Kelley J. Main, and Scott A. Roy. 2000. Meta-stereotype activation: evidence from indirect measures for specific evaluative concerns experienced by members of dominant groups in intergroup interaction. *Journal of Personality and Social Psychology*, 78(4):690–707.

Cheryl J. Wakslak, Pamela K. Smith, and Albert Han. 2014. Using abstract language signals power. *Journal of Personality and Social Psychology*, 107(10):41–55.

Lien Wille and Eva Derous. 2017. Getting the words right: When wording of job ads affects ethnic minorities' application decisions. *Management Communication Quarterly*, 31(4):533–558.

Eva Zschirnt and Didier Ruedin. 2016. Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. *Journal of Ethnic and Migration Studies*, 42(7):1115–1134.

A   Appendix Adaptions of LCM to job advertisements

The Linguistic Category Model (Semin and Fiedler, 1991) describes ways of communication in the interpersonal domain that covers social interaction between people. The same interpersonal event can be expressed in various ways. For example, a fight can be described behaviorally (on a physical level) such as [subj] *kicks* [obj] or dispositionally (on a mental level) such as [subj] *despises* [obj].

Job advertisements, however, do not exactly fall into the category of direct interpersonal communication that is covered by the LCM as presented by Semin and Fiedler (1991). In the advertisement texts, the applicant is most of the time the subject and the verbs relate them either to another person or group of persons (e.g. *Je spoort je collega's aan* English: 'You encourage your colleagues'), an action (e.g. *Je presenteert je bevindingen* English: 'You present your findings'), or an object (e.g. *Je brengt de krant rond* English: 'You deliver the newspaper'). This means that not all definitions of the categories as defined in the LCM match precisely with the intent of this task. Therefore, the model had to be adapted to the new domain of job advertisements. Adapting the Linguistic Category Model to the context of job advertisements, the following labels were obtained:

- **Descriptive Action Verb was given the label "Act"**
DAV was translated to "Act" and described as a single action that can be easily visualized and usually started and completed in a few hours. It is distinguishable with a physically invariant feature.

    Example: *knippen van vlakke platen* English: 'cutting of flat sheets'. Cutting is based on a verb, describing an action with beginning and end, with a physically invariant feature (the action is done by hand). This is the most concrete type of phrasing.

- **Interpretive Action Verb was given the label "Process"**
IAV was translated to "Process", which is a series of acts or one that can be visualized and/or interpreted in multiple ways. The process is an action that is not distinguished by a physically invariant feature. It has a beginning and end but may take more time (up to days, weeks or months) to complete than an Act.

    Example: *aansturen van vijf medewerkers, werkvoorbereiding / calculatie doorvoeren en inmeten* English: 'managing five employees, carrying out work preparation / entering calculations and measuring'. Managing, entering, and measuring are all verbs describing actions with no positive or negative valence, with a beginning and end, but without physically invariant feature (managing can be done by pointing/talking/writing, etc.).

    Example: *Kortom: je weet klantbehoeftes door te vertalen naar oplossingen en een brug te slaan* English: 'In short: you know how to translate customer needs into solutions and bridge the gap'. To translate and bridge a gap are actions that generally need

some amount of interpretation to be understood in context. They are not completely self-explanatory. Translating in this sense is not translation between two languages, and similarly bridging a gap does not mean to physically build a bridge brick by brick. It rather implies a process of finding solutions for problems. Both consist of a combination of more concrete actions.

- **State Verb was given the label "Attitude"**
  SV is called an "Attitude" and should refer to a psychological enduring state, a way of 'being' that is constant over time with a verb as basis. That is, in the context of job ads, a stable way of thinking or feeling. These states cannot be objectively verified.

  Example: *Daarin denk je vanuit concepten* English: 'Therein, you think in concepts'. A way of thinking is not an action but rather a way of 'being' that is stable over time.

  Example: *Je hebt een instelling van wat kan wel i.p.v. wat kan niet* English: 'You have an attitude that looks at what is possible instead of what is not'. This describes a psychological state showing a consequent reaction to being faced with a problem.

- **State Action Verb was given the label "Attitude + action"**
  SAV is called an "Attitude + action" and refers to a psychological enduring state just like a SV, as a result of an action.

  Example: *Je krijgt er energie van op 5 borden tegelijk te schaken* English: 'You get energized from playing chess on 5 boards simultaneously'. Getting energized is a resulting psychological state of performing the action which is playing chess on 5 boards - a metaphor for multitasking.

- **Adjective / Noun / Adverb was given the label "Quality"**
  The label given to the ADJ/NOUN/ADV class is "Quality", because these phrases should describe what the ideal employee is like, thus, what qualities the job advertisement mentions that the person should have. This could be personality traits, skills, or qualifications.

  Example: *Functie eisen: je hebt uitstekende analytische en communicatieve*

vaardigheden English: 'Job requirements: you have excellent analytical and communicative skills'. An adjective like "excellent" plus a noun like "skills" that describe someone's stable qualities without specifying what kind of behavior contributes to this makes that this is the most abstract type of phrasing. Qualities of the company, actions, or objects should not be annotated, as those are irrelevant for the research question.

"Quality" is further divided into the sub-labels "Innate quality" and "Learned quality". Where Semin and Fiedler (1991) only discusses innate qualities like 'honest' and 'impulsive', job advertisements contain many required qualities such as *Je beheerst de Engelse taal* English: 'You master the English language', *Je hebt een rijbewijs* English: 'You have a drivers license', or *Je hebt aantoonbare kennis van Excel* English: 'You have demonstrable knowledge of Excel' which are skills not acquired by nature but by active learning or training. This is an important distinction to make because the innate qualities can not be validated easily, while the learned ones can be validated with a certificate or test. Besides, the innate qualities tell more about qualities that play a role in the interpersonal domain whereas the learned qualities generally do not.