Word(s) and Object(s): Grounded Language Learning In Information Retrieval

Federico Bianchi*
Bocconi University
Milano, Italy
f.bianchi@unibocconi.it

Ciro Greco
Coveo AI Labs
New York, United States
cgreco@coveo.com

Jacopo Tagliabue
Coveo AI Labs
New York, United States
jtagliabue@coveo.com

Abstract

We present a grounded language model for Information Retrieval, that learns lexical *and* compositional meaning for search queries from dense representations of objects; in our case, target entities are products, modeled as low-dimensional embeddings trained over behavioural data from a e-commerce website. Crucially, the proposed semantics exhibits compositional virtues but it is still fully learnable without explicit labelling: our domain of reference, denotation and composition are all learned from user data only. We benchmark the grounded model against SOTA intra-textual models (such as word2vec and BERT), and we show that it provides more accuracy and better generalizations.

1 Introduction

That words refer to objects is a platitude hardly in need of defense: scholarly work in psycholinguistics [Xu and Tenenbaum, 2000], formal and informal semantics [Montague, 1974, Chierchia and McConnell-Ginet, 2000], philosophy of language – such as the seminal [Quine, 1960], literally titled *Word and Object* – assume that at least part of linguistic meaning can be represented as some sort of mapping between language and extra-linguistic entities; in the same vein, theoretical and simulation-based work [Fitzgerald and Tagliabue, 2020, Lazaridou et al., 2018] on the emergence of language subscribes to the same idea: *language games* in Wittgenstein [1953] involve mapping sounds to salient objects; *signaling games* – whether classically or evolutionary construed [Lewis, 1969, Skyrms, 1997] – take place with agents communicating events in their surroundings. In contrast, most SOTA models in NLP are only *intra-textual*. Distributional semantics (*DS*) – such as standard and contextual word embeddings [Mikolov et al., 2013, Peters et al., 2018] – learns (representations of) meaning from patterns of co-occurrence in big corpora, with no reference to non-linguistics entities. As far BERT [Devlin et al., 2019] knows, language is all there is and the world has no bearing in extra-linguistic terms.

In *this* work, we propose a new language model for Information Retrieval, in which unsupervised learning is combined with a language-independent domain of reference. Our semantics is grounded in object representations and compositionality is attained through those representations [Heim and Kratzer, 1998]. Crucially, in our proposal the extra-linguistic reference domain is fully learnable: in particular, we propose to learn our reference domain, denotation function and compositional semantics from scratch, using only behavioural data and search engine interactions from an online shop. Our grounded model has the main purpose to capture one of the most fundamental intuitions about meaning: namely, language expressions are *about something* and that something is often *non-linguistic*.

*Corresponding author. All authors contributed equally and are listed alphabetically.

NeurIPS 2020 Workshop on Object Representations for Learning and Reasoning, Vancouver, Canada.

Domain of reference: we learn neural representations for our domain of objects – in our dataset, products in an online shop – with prod2vec [Grbovic et al., 2015]; we do *not* leverage manually built discrete structures [Lu et al., 2018], nor embeddings of these structures [Hamilton et al., 2018]. Crucially, building the domain of reference does not require any linguistic input and it is independently justified in the target domain [Tagliabue et al., 2020a,b].

Lexical denotation: we use the natural interactions between shoppers and a search engine to learn a noisy denotation for the basic lexicon: the denotation of "shoes" is given by shoppers clicking on products which are in fact *shoes* after issuing the search query "shoes". In other words, we reinterpret game-based models of language acquisition [Lewis, 1969, Tomasello, 2003] in the cooperative context of information retrieval [Tagliabue and Cohn-Gordon, 2019]: by substituting discrete sets with point clouds, the meaning of a word such as "shoes" can still be represented as a function, while its extra-textual denotation is now a *DeepSet* [Cotter et al., 2018].

Compositional semantics: we replace a discrete formal semantics of noun phrases with functions learned over DeepSets, and we test the generalization capability of the model on zero-shot inference – once we have learned "Nike shoes", can we reliably predict the composition of "Adidas shorts" as well?

1.1 Contributions

We summarize our main contributions as follows: first, we introduce a novel hybrid semantics grounded in object representations, and we build a compositional language from the lexical meaning and the topology of the underlying dense space. Albeit domain specific, our language is based on 26k products from a real e-commerce website, and as such it is significantly richer than languages from agent-based models [Słowik et al., 2020]; second, our model does not assume any available set of entities, nor a denotation function. This represents a major departure from standard work on "grounded language", where manually constructed data sources have built-in compositional properties [Krishna et al., 2016], and work in multi-modal language understanding [Antol et al., 2015, Mogadala et al., 2018, Günther et al., 2020], where compositionality is learned from explicit image tagging; in other words, unsupervised representations are used to show the non-obvious fact that a dense domain supports compositionality, especially when compared to graph-based representations [Hamilton et al., 2018], where discrete objects are built-in by design. Finally, our focus on Information Retrieval is put forward as a methodological strength, as search engine dynamics is a sweet spot between the simplifying assumptions of toy research models and the daunting task of providing a general account of, say, the semantics of English: products are a domain rich enough to provide a wealth of data and, possibly, practical applications, while at the same time sufficiently self-contained to be realistically mastered without human supervision. Philosophically, our work is sympathetic with the recent rediscovery of extra-textual elements [Bisk et al., 2020] and with broader reflections on "meaning" in the community [Bender and Koller, 2020]: if we "cannot learn language from the radio", perhaps we can partially learn it from a search bar.

2 Methods

Following our informal exposition in Section 1, we distinguish three components, which are learned separately in a sequence: learning a language-independent object domain (namely, the product space), learning noisy denotation from search logs and finally learning functional composition. While only the first model (prod2vec) is completely unsupervised, it is important to remember that the other learning procedures are only weakly supervised, as the labelling is obtained by exploiting an existing user-machine dynamics to provide noisy labels (i.e. no human curation was necessary at any stage of the training process).

Learning a product space. Dense object representations are known in e-commerce as product embeddings [Grbovic et al., 2015]. Following industry standards, a *prod2vec* model is trained as a skip-gram model over browsing sessions [Mu et al., 2018]: prod2vec models are word2vec models in which words in a sentence are replaced by products in a session[2] – similar to what happens with word2vec, related products will end up closer in the embedding space.

[2]For this study, we pick $d = 24$ as our vector size, optimizing hyperparameters as in Bianchi et al. [2020].

Learning lexical denotation. We interpret the click on products in the search result page, after a query is issued, as a noisy "pointing" signal, i.e., a map between text ("shoes") and the target domain (a portion of the product space). We represent the meaning of search queries through an order-invariant operation over the embeddings of the clicked products (average pooling weighted by empirical frequencies, similar to Yu et al. [2020]). We refer to this representation as a *DeepSet* [Cotter et al., 2018].

Learning functional composition. Our functional composition will come from the composition of DeepSet representations, where we want to learn a function $f : DeepSet \times DeepSet \rightarrow DeepSet$. We address functional composition by means of two models from the relevant literature [Hartung et al., 2017]: one, *Additive Compositional Model* (**ADM**), sums vectors together to build the final DeepSet representation. The second model is instead a *Matrix Compositional Model* (**MDM**): given in input two DeepSets (for example, one for "Nike" and one for "shoes") the function we learn as the form $Mv + Nu$, where the interaction between the two vectors is mediated through the learning of two matrices, M and N. Since the output of these processes is always a DeepSet, both models can be recursively composed, given the form of the function f.

3 Experiments

Data. To the best of our knowledge, no dataset with session-level shopping data and matching search engine behavior is openly available [3]. To test our model, we obtained catalog data, search logs and detailed behavioral data (anonymized product interactions) from a partnering online shop, *Shop X*. *Shop X* is an Italian e-commerce website in the sport apparel vertical, with revenues between 20 and 100 million USD/year. Browsing and search data are sampled from the summer of 2019, for a total, after pre-processing, of $722,479$ sessions, $26,057$ product embeddings and $37,689$ distinct query strings. Due to the power-law distribution of search queries, one-word queries are the majority of the dataset (60%); to compensate for sparsity in our dataset we perform data augmentation for rare complex queries (e.g. "Nike running shoes"): after we send a query to the existing search API to get a result set, we simulate $n = 500$ clicks by drawing products from the set, with probability proportional to their popularity [4]. The final dataset consists of 818 "brand + sortal"[5] queries – e.g. "Nike shoes" –, 104 "sortal + activity"[6] queries – "running shoes" –, and 47 "sortal + gender" queries – "women shoes"; our testing data consists of 521 "sortal + activity + brand" (SAB) triples, 157 "sortal + activity + gender" (SAG) triples, 406 "sortal + activity + gender + brand" (SAGB) quadruples.[7]

Tasks and Metrics. We evaluate models in relation to their ability to predict the DeepSet for a given query – given "shoes" and "Nike", the model is trained to predict the DeepSet representation of "Nike shoes". To evaluate this, the nearest neighbors of the predicted DeepSet are compared to the ones of the target DeepSet from the dataset, using three well-known evaluation ranking metrics: *nDCG*, *Hit Rate@10* (HITS) and *Jaccard* [Mitra and Craswell, 2018, Vasile et al., 2016, Jaccard, 1912]. We focus on two tasks: leave-one-brand-out (**LOBO**) and zero-shot (**ZT**). In **LOBO**, we train models over the "sortal brand" queries but we exclude from training a specific brand (e.g., "Nike"); in the test phase, we ask the models to predict the DeepSet for a seen sortal and an unseen brand. For **ZT** we train models over queries with two terms ("sortal + brand", "sortal + activity" and "sortal + gender") and see how well our semantics generalizes to compositions like "sortal + activity + brand"; the complex queries that we used at test time are new and unseen.

[3]Following a recent work from *Coveo Labs* involving time-series classification [Requena et al., 2020], we plan the release of anonymized e-commerce datasets to further research activity in IR and NLP. Contact the authors for further information.

[4]Since the only products (real or simulated) users can click are the ones returned by the search API, query representation may in theory be biased by the peculiarities of the engine. In practice, preliminary tests confirmed that the embedding quality is stable even when a sophisticated search engine is replaced by Boolean queries over TF-IDF vectors, suggesting that any bias of this sort is likely to be very small and not important for the quality of the compositional semantics.

[5]"Sortal" refers to a *type* of object: *shoes* and *polo* are sortals, while *black* and *Nike* are not.

[6]"Activity" is the sport activity for a product, e.g. *tennis* for a racquet.

[7]While not massive, dataset size for our compositional tests is in line with intra-textual studies on compositionality [Baroni and Zamparelli, 2010, Rubinstein et al., 2015]; moreover, the lexical atoms in our study reflect a real-world distribution that is independently generated, and not frequency on general English corpora.

Models. We benchmark four models on our two tasks: the compositional models with our DeepSet semantics: **ADM** and **MDM**; as intra-textual baselines for meaning representation, we pick contextual and non-contextual embeddings from state-of-the-art *DS* models: BERT (**UM**) (the Umberto model[8]) and Word2Vec (**W2V**), trained on product descriptions from *Shop X* catalog. For **UM**, we extract the 768 dimensional representation from the [CLS] embedding of the 12th layer of the query and learn a linear projection to the product-space (essentially, training to predict the DeepSet representation from text). The generalization to different and longer queries for **UM** comes from the embeddings of the queries themselves. Instead, for **W2V**, we learn a compositional function that concatenates the two input DeepSets, projects them to 24 dimensions, pass them through a Rectified Linear Unit, and finally project them to the product space.[9] We run every model 15 times and report average results; RMSProp is the chosen optimizer, with a batch size of 200, 20% of the training set as validation set and early stopping with $patience = 10$.

Table 1: Test scores on **LOBO**, best in bold.

LOBO	**ADM**	**MDM**	**UM**	**W2V**
nDCG	0.1228	**0.187**	0.0024	0.0098
HITS	0.1821	**0.2993**	0.0030	0.0138
Jaccard	0.0713	**0.1175**	0.0009	0.0052

Table 2: Test scores on **ZT**, best in bold.

ZT	**ADM**	**MDM**	**UM**	**W2V**
SAB				
nDCG	0.0810	**0.0988**	0.0312	0.0064
HITS	0.0616	**0.0736**	0.0251	0.0052
Jaccard	0.0348	**0.0383**	0.0113	0.0023
SAG				
nDCG	**0.0221**	0.0078	0.019	0.0005
HITS	**0.0210**	0.0061	0.0157	0.0006
Jaccard	**0.0083**	0.0022	0.0052	0.0001
SAGB				
nDCG	0.0332	**0.0375**	0.0124	0.0059
HITS	0.0295	**0.0305**	0.0102	0.0039
Jaccard	0.0162	**0.0163**	0.0044	0.0019

Results. Table 1 shows the chosen models on **LOBO**, with **MDM** beating **ADM** and outperforming all intra-textual models. Table 2 reports performance for different complex query types in the zero-shot inference task: **MDM** outperforms **ADM** and the intra-textual models for *SAB* and *SAGB*, but all models suffer from *gender* sparsity in that case; unsurprisingly, the best model for *SAG* is **ADM**, that is, the only one that does not come with an implicit bias from the training. Generally speaking, in all cases, grounded models outperform intra-textual models, often by a wide margin. The quantitative evaluations were confirmed by manually inspecting nearest neighbors for predicting DeepSets in the **LOBO** setting – as an example, **MDM** predicts for "Nike shoes" a DeepSet that has (correctly) all *shoes* as neighbors in the space, while, for the same query, **UM** suggests *shorts* as answer.

4 Conclusions and Future Work

We presented a grounded language model for Information Retrieval, in which all the important pieces – domain, denotation, composition – are learned from behavioral data. By grounding meaning in (a representation of) objects and their properties, the proposed semantics can be learned "bottom-up" like distributional models, but can generalize to unseen examples, like traditional referential models: the implicit, dense structure of the domain (e.g. the relative position in the space of *Nike* products and *shoes*) underpins the explicit, discrete structure of queries picking objects in that domain (e.g. "Nike shoes") – in other words, compositionality is an emergent phenomenon. While encouraging, our results are still preliminary: first, we plan on extending the language model, starting with Boolean operators ("shoes *NOT* Nike"); second, we plan to improve our representational capabilities with regard to products, either through symbolic knowledge or more discerning embedding strategies; third, we wish to explore transformer-based architecture [Lee et al., 2018] as an alternative way to produce set-like representations.

[8] `https://huggingface.co/Musixmatch/umberto-commoncrawl-cased-v1`

[9] First results with the same structure as ADM and MDM showed very low performances, thus we made the architecture more complex and non-linear.

References

Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, and Devi Parikh. VQA: Visual Question Answering. In *International Conference on Computer Vision (ICCV)*, 2015.

Marco Baroni and Roberto Zamparelli. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1183–1193, Cambridge, MA, October 2010. Association for Computational Linguistics. URL `https://www.aclweb.org/anthology/D10-1115`.

Emily M. Bender and Alexander Koller. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.463. URL `https://www.aclweb.org/anthology/2020.acl-main.463`.

Federico Bianchi, Jacopo Tagliabue, Bingqing Yu, Luca Bigon, and Ciro Greco. Fantastic embeddings and how to align them: Zero-shot inference in a multi-shop scenario. In *Proceedings of the SIGIR 2020 eCom workshop, July 2020, Virtual Event, published at http://ceur-ws.org (to appear)*, 2020. URL `https://arxiv.org/abs/2007.14906`.

Yonatan Bisk, Ari Holtzman, Jesse Thomason, Jacob Andreas, Yoshua Bengio, Joyce Chai, Mirella Lapata, Angeliki Lazaridou, Jonathan May, Aleksandr Nisnevich, Nicolas Pinto, and Joseph Turian. Experience grounds language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8718–8735, Online, November 2020. Association for Computational Linguistics. URL `https://www.aclweb.org/anthology/2020.emnlp-main.703`.

Gennaro Chierchia and Sally McConnell-Ginet. *Meaning and Grammar (2nd Ed.): An Introduction to Semantics*. MIT Press, Cambridge, MA, USA, 2000. ISBN 026253164X.

Andrew Cotter, Maya R. Gupta, Heinrich Jiang, James Muller, Taman Narayan, Serena Wang, and Tao Zhu. Interpretable set functions. *ArXiv*, abs/1806.00050, 2018.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL `https://www.aclweb.org/anthology/N19-1423`.

Nicole J M Fitzgerald and J. Tagliabue. On the plurality of graphs. *ArXiv*, abs/2008.00920, 2020.

Mihajlo Grbovic, Vladan Radosavljevic, Nemanja Djuric, Narayan Bhamidipati, Jaikit Savla, Varun Bhagwan, and Doug Sharp. E-commerce in your inbox: Product recommendations at scale. In *Proceedings of KDD '15*, August 2015. doi: http://dx.doi.org/10.1145/2783258.2788627. URL `https://astro.temple.edu/~tuc17157/pdfs/grbovic2015kddB.pdf`.

Fritz Günther, Marco Alessandro Petilli, and Marco Marelli. Semantic transparency is not invisibility: A computational model of perceptually-grounded conceptual combination in word processing. *Journal of Memory and Language*, 112:104104, 2020.

William L. Hamilton, Payal Bajaj, Marinka Zitnik, Dan Jurafsky, and Jure Leskovec. Embedding logical queries on knowledge graphs. In *NeurIPS*, 2018.

Matthias Hartung, Fabian Kaupmann, Soufian Jebbara, and Philipp Cimiano. Learning compositionality functions on word embeddings for modelling attribute meaning in adjective-noun phrases. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 54–64, 2017.

Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Wiley-Blackwell, 1998. ISBN 0631197125, 0631197133, 9780631197133.

Paul Jaccard. The distribution of the flora in the alpine zone.1. *New Phytologist*, 11(2):37–50, 1912. doi: 10.1111/j.1469-8137.1912.tb05611.x. URL `https://nph.onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8137.1912.tb05611.x`.

Ranjay Krishna, Yuke Zhu, Oliver Groth, Justin Johnson, Kenji Hata, Joshua Kravitz, Stephanie Chen, Yannis Kalantidis, Li-Jia Li, David A. Shamma, Michael S. Bernstein, and Li Fei-Fei. Visual genome: Connecting language and vision using crowdsourced dense image annotations. *International Journal of Computer Vision*, 123:32–73, 2016.

A. Lazaridou, K. Hermann, K. Tuyls, and Stephen Clark. Emergence of linguistic communication from referential games with symbolic and pixel input. *ArXiv*, abs/1804.03984, 2018.

Juho Lee, Y. Lee, Jungtaek Kim, Adam R. Kosiorek, S. Choi, and Y. Teh. Set transformer. *ArXiv*, abs/1810.00825, 2018.

David Lewis. *Convention: A Philosophical Study*. Harvard University Press, Cambridge, 1969.

Di Lu, Spencer Whitehead, Lifu Huang, Heng Ji, and Shih-Fu Chang. Entity-aware image caption generation. In *EMNLP*, 2018.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 3111–3119, Red Hook, NY, USA, 2013. Curran Associates Inc.

Bhaskar Mitra and Nick Craswell. An introduction to neural information retrieval. *Foundations and Trends® in Information Retrieval*, 13(1):1–126, December 2018. URL `https://www.microsoft.com/en-us/research/publication/introduction-neural-information-retrieval/`.

Aditya Mogadala, Umanga Bista, Lexing Xie, and Achim Rettinger. Knowledge guided attention and inference for describing images containing unseen objects. In *ESWC*, 2018.

Richard Montague. Universal grammar. In Richmond H. Thomason, editor, *Formal Philosophy: Selected Papers of Richard Montague*, number 222–247. Yale University Press, New Haven, London, 1974.

Cun Mu, Guang Yang, and Zheng Yan. Revisiting skip-gram negative sampling model with regularization. *CoRR*, abs/1804.00306, 2018. URL `http://arxiv.org/abs/1804.00306`.

Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1202. URL `https://www.aclweb.org/anthology/N18-1202`.

W. V. O. Quine. *Word & Object*. MIT Press, 1960.

Borja Requena, Giovanni Cassani, Jacopo Tagliabue, Ciro Greco, and Lucas Lacasa. Shopper intent prediction from clickstream e-commerce data with minimal browsing information. *Scientific Reports*, 10:2045–2322, 2020. doi: 10.1038/s41598-020-73622-y.

Dana Rubinstein, Effi Levi, Roy Schwartz, and Ari Rappoport. How well do distributional models capture different types of semantic knowledge? In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 726–730, Beijing, China, July 2015. Association for Computational Linguistics. doi: 10.3115/v1/P15-2119. URL `https://www.aclweb.org/anthology/P15-2119`.

Brian Skyrms. Evolution of the social contract. *Philosophy*, 72(282):604–606, 1997.

Agnieszka Słowik, Abhinav Gupta, William L. Hamilton, Mateja Jamnik, Sean B. Holden, and Christopher Joseph Pal. Exploring structural inductive biases in emergent communication. *ArXiv*, abs/2002.01335, 2020.

Jacopo Tagliabue and Reuben Cohn-Gordon. Lexical learning as an online optimal experiment: Building efficient search engines through human-machine collaboration. *ArXiv*, abs/1910.14164, 2019.

Jacopo Tagliabue, Bingqing Yu, and Marie Beaulieu. How to grow a (product) tree: Personalized category suggestions for eCommerce type-ahead. In *Proceedings of The 3rd Workshop on e-Commerce and NLP*, pages 7–18, Seattle, WA, USA, July 2020a. Association for Computational Linguistics.

Jacopo Tagliabue, Bingqing Yu, and Federico Bianchi. The embeddings that came in from the cold: Improving vectors for new and rare products with content-based inference. In *Fourteenth ACM Conference on Recommender Systems*, RecSys '20, page 577–578, New York, NY, USA, 2020b. Association for Computing Machinery. ISBN 9781450375832. doi: 10.1145/3383313.3411477. URL `https://doi.org/10.1145/3383313.3411477`.

Michael Tomasello. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, Cambridge, MA, 2003.

Flavian Vasile, Elena Smirnova, and Alexis Conneau. Meta-prod2vec: Product embeddings using side-information for recommendation. *Proceedings of the 10th ACM Conference on Recommender Systems*, 2016.

Ludwig Wittgenstein. *Philosophical Investigations*. Basil Blackwell, Oxford, 1953. ISBN 0631119000.

Fei Xu and Joshua B. Tenenbaum. Word learning as bayesian inference. In *In Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, pages 517–522. Erlbaum, 2000.

Bingqing Yu, Jacopo Tagliabue, Federico Bianchi, and Ciro Greco. An image is worth a thousand features: Scalable product representations for in-session type-ahead personalization. In *Companion Proceedings of the Web Conference*, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 978-1-4503-7024-0/20/04. doi: 10.1145/3366424.3386198.