

# Grounding “grounding”: How is grounding used within various AI conferences?

Eric Huang

University of Waterloo  
e48huang@uwaterloo.ca

## Abstract

Terminology used within linguistics and AI conferences tend to be overused, leading to ambiguous meaning and difficulty navigating new papers. This paper will elucidate the various senses of the word “grounding” through qualitative analysis and deeper quantitative analysis of its various uses. This work will showcase how “grounding” is an overloaded term and guide users to understand how to more easily decipher and understand papers using the term. All code can be found at [https://git.uwaterloo.ca/e48huang/cs-784/-/tree/final\\_project/final\\_project?ref\\_type=heads](https://git.uwaterloo.ca/e48huang/cs-784/-/tree/final_project/final_project?ref_type=heads) where a University of Waterloo account is required.

## 1 Introduction

Many conferences centering around Artificial Intelligence have existed for many decades, evolving over time on the types of problems that they tackle. While these problems change over time, so do the terminology, which have a tendency to evolve semantically, leading to overloaded terms. One such term is “grounding”, the idea that one wishes to ensure that there is understanding or a common ground (Nakano et al., 2003). While this term seems simple, it is used in many various contexts, all of which requires different datasets, methods and metrics to evaluate, while being applied in different settings.

To better understand the term “grounding” and its usage, we perform both quantitative analysis and qualitative analysis. This paper explores the “Seed42Lab/AI-paper-crawl” HuggingFace dataset (Forty-Two AI Lab) which collects full-text papers from 11 different conferences spanning from the first year of the conference to 2024. To first select different senses of the word “grounding”, we perform preliminary quantitative analysis to filter for papers to further investigate. From these selected

Conference	Paper Count
AAAI	772
ACL	632
CVPR	862
ECCV	511
EMNLP	575
ICCV	341
ICLR	360
ICML	360
IJCAI	654
NAACL	226
NeurIPS	654

Table 1: Counts of unique papers with “grounding” by conference found in the corpora.

papers, we identify 9 related but distinct meanings of the word “grounding”. We perform some literature review to understand how these different senses are understood, from its various datasets, methods, metrics and applications. Finally, for each of these word senses, we investigate how they have evolved over time.

## 2 Paper Selection

A simple search over the number of papers which have the term “grounding” quickly shows that it is infeasible to cover all possible instances. For example, the Association for Computational Linguistics (ACL) alone has 632 unique papers that have an instance of “grounding” (see Table 1). While not all these instances are due to the paper itself being related to grounding, as they can simply include it within its bibliography, they are indicative that some filtering is necessary.

To filter through these papers, we propose a method which selects the most relevant papers within a conference to the word “grounding”. We take a naive approach where we select the top 10% of papers with the word “grounding”. We deter-

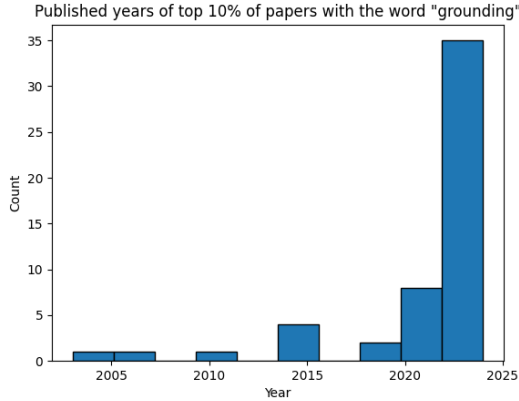


Figure 1: Count of selected conference papers per year

mine which papers are more important to “grounding” based on the word frequency, if “grounding” appears more often compared to other words within a paper, then it should be more relevant. This process (while ensuring uniqueness across conferences) resulted in 46 curated papers<sup>1</sup>, spanning from the years 2000 to 2024, shown in the following figure. While this selection of papers may not cover the breadth of senses of grounding might entail, as it misses on papers from the 1980’s to 2000’s, it does cover the most commonly used senses of the word.

### 3 Grounding “grounding”

In this section we cover the different senses of the word “grounding” found in the 46 papers covered in the previous section. We will explore the common methods, datasets, applications and metrics revolving around each word. We will continue to elucidate the meanings of the word and its trend within the next section.

#### 3.1 Visual Grounding

Visual grounding, also known as image, phrase or referring expression grounding (Xiao et al., 2024; Li et al., 2024) refers to the challenge of trying to localize specific regions within an image based on some textual description. Traditionally, this involved finding the phrase’s referring region by predicting a bounding box around said region. As time has gone on however, there has been more and more types of challenges that one could tackle within visual grounding (Xiao et al., 2024). From the 46

<sup>1</sup>[https://git.uwaterloo.ca/e48huang/cs-784/-/blob/e09a1c22c0de7e331ca16109a5f32b226dc6d9c5/final\\_project/grounding\\_top\\_p.txt](https://git.uwaterloo.ca/e48huang/cs-784/-/blob/e09a1c22c0de7e331ca16109a5f32b226dc6d9c5/final_project/grounding_top_p.txt)

papers we filtered for, we have discovered the following subcategories involved in visual grounding:

1. Classical visual grounding
2. Gigapixel level grounding. TODO: add citations
3. Gigapixel level grounding

#### 3.2 Spatio-temporal Grounding

#### 3.3 3D Grounding

#### 3.4 Video and Audio Grounding

#### 3.5 Markov Logic Networks Grounding

#### 3.6 Face-to-Face Grounding

#### 3.7 Action Grounding

#### 3.8 Relational Grounding

### 4 How has “grounding” evolved over time?

In this section, we will take each previous section’s meaning of “grounding” and build an understanding of its trends through observing the co-occurrence over time with other key words for each specific sense.

#### 4.1 Visual Grounding

TODO: Also explore why there are so many different ways to mean visual grounding. Also explore which datasets are most popular within these papers.

#### 4.2 Spatio-temporal Grounding

#### 4.3 3D Grounding

#### 4.4 Video and Audio Grounding

#### 4.5 Markov Logic Networks Grounding

#### 4.6 Face-to-Face Grounding

#### 4.7 Action Grounding

#### 4.8 Relational Grounding

### 5 Engines

To produce a PDF file, pdfL<sup>A</sup>T<sub>E</sub>X is strongly recommended (over original L<sup>A</sup>T<sub>E</sub>X plus dvips+ps2pdf or dvipdf). The style file acl.sty can also be used with luaL<sup>A</sup>T<sub>E</sub>X and XeL<sup>A</sup>T<sub>E</sub>X, which are especially suitable for text in non-Latin scripts. The file acl\_lualatex.tex in this repository provides an example of how to use acl.sty with either luaL<sup>A</sup>T<sub>E</sub>X or XeL<sup>A</sup>T<sub>E</sub>X.

## 6 Preamble

The first line of the file must be

```
\documentclass[11pt]{article}
```

To load the style file in the review version:

```
\usepackage[review]{acl}
```

For the final version, omit the review option:

```
\usepackage{acl}
```

To use Times Roman, put the following in the preamble:

```
\usepackage{times}
```

(Alternatives like `txfonts` or `newtx` are also acceptable.)

Please see the  $\LaTeX$  source of this document for comments on other packages that may be useful.

Set the title and author using `\title` and `\author`. Within the author list, format multiple authors using `\and` and `\And` and `\AND`; please see the  $\LaTeX$  source for examples.

By default, the box containing the title and author names is set to the minimum of 5 cm. If you need more space, include the following in the preamble:

```
\setlength\titlebox{<dim>}
```

where `<dim>` is replaced with a length. Do not set this length smaller than 5 cm.

## 7 Document Body

### 7.1 Footnotes

Footnotes are inserted with the `\footnote` command.<sup>2</sup>

### 7.2 Tables and figures

See Table 2 for an example of a table and its caption.

**Do not override the default caption sizes.**

As much as possible, fonts in figures should conform to the document fonts. See Figure 2 for an example of a figure and its caption.

Using the `graphicx` package `graphics` files can be included within figure environment at an appropriate point within the text. The `graphicx` package supports various optional arguments to control the appearance of the figure. You must include it explicitly in the  $\LaTeX$  preamble (after the `\documentclass` declaration and before `\begin{document}`) using `\usepackage{graphicx}`.

<sup>2</sup>This is a footnote.

Command	Output	Command	Output
<code>\`a</code>	ä	<code>\c c</code>	ç
<code>\^e</code>	ê	<code>\u g</code>	ğ
<code>\`i</code>	ì	<code>\l</code>	ł
<code>\.I</code>	İ	<code>\~n</code>	ñ
<code>\o</code>	ø	<code>\H o</code>	ő
<code>\'u</code>	ú	<code>\v r</code>	ř
<code>\aa</code>	å	<code>\ss</code>	ß

Table 2: Example commands for accented characters, to be used in, e.g., Bib $\TeX$  entries.



Figure 2: A figure with a caption that runs for more than one line. Example image is usually available through the `mwe` package without even mentioning it in the preamble.

### 7.3 Hyperlinks

Users of older versions of  $\LaTeX$  may encounter the following error during compilation:

```
\pdfendlink ended up in different nest-
ing level than \pdfstartlink.
```

This happens when `pdf $\LaTeX$`  is used and a citation splits across a page boundary. The best way to fix this is to upgrade  $\LaTeX$  to 2018-12-01 or later.

### 7.4 Citations

Table 3 shows the syntax supported by the style files. We encourage you to use the `natbib` styles. You can use the command `\citet` (cite in text) to get “author (year)” citations, like this citation to a paper by ?. You can use the command `\citep` (cite in parentheses) to get “(author, year)” citations (?). You can use the command `\citealp` (alternative cite without parentheses) to get “author, year” citations, which is useful for using citations within parentheses (e.g. ?).

A possessive citation can be made with the command `\citeposs`. This is not a standard `natbib` command, so it is generally not compatible with other style files.

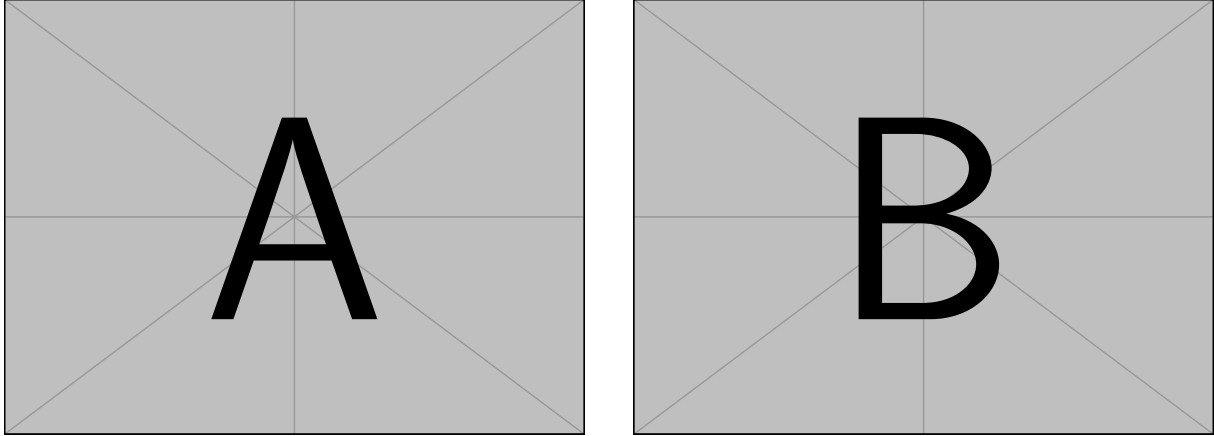


Figure 3: A minimal working example to demonstrate how to place two images side-by-side.

Output	natbib command	ACL only command
(?)	<code>\citep</code>	
?	<code>\citealp</code>	
?	<code>\citett</code>	
(?)	<code>\citeyearpar</code>	
?’s (?)		<code>\citeposs</code>

Table 3: Citation commands supported by the style file. The style is based on the natbib package and supports all natbib citation commands. It also supports commands defined in previous ACL style files for compatibility.

## 7.5 References

The  $\text{\LaTeX}$  and Bib $\text{\TeX}$  style files provided roughly follow the American Psychological Association format. If your own bib file is named `custom.bib`, then placing the following before any appendices in your  $\text{\LaTeX}$  file will generate the references section for you:

```
\bibliography{custom}
```

You can obtain the complete ACL Anthology as a Bib $\text{\TeX}$  file from <https://aclweb.org/anthology/anthology.bib.gz>. To include both the Anthology and your own .bib file, use the following instead of the above.

```
\bibliography{anthology,custom}
```

Please see Section 8 for information on preparing Bib $\text{\TeX}$  files.

## 7.6 Equations

An example equation is shown below:

$$A = \pi r^2 \quad (1)$$

Labels for equation numbers, sections, subsections, figures and tables are all defined with the `\label{label}` command and cross references to them are made with the `\ref{label}` command.

This an example cross-reference to Equation 1.

## 7.7 Appendices

Use `\appendix` before any appendix section to switch the section numbering over to letters. See Appendix A for an example.

## 8 Bib $\text{\TeX}$ Files

Unicode cannot be used in Bib $\text{\TeX}$  entries, and some ways of typing special characters can disrupt Bib $\text{\TeX}$ ’s alphabetization. The recommended way of typing special characters is shown in Table 2.

Please ensure that Bib $\text{\TeX}$  records contain DOIs or URLs when possible, and for all the ACL materials that you reference. Use the `doi` field for DOIs and the `url` field for URLs. If a Bib $\text{\TeX}$  entry has a URL or DOI field, the paper title in the references section will appear as a hyperlink to the paper, using the `hyperref`  $\text{\LaTeX}$  package.

## Limitations

Since December 2023, a "Limitations" section has been required for all papers submitted to ACL Rolling Review (ARR). This section should be placed at the end of the paper, before the references. The "Limitations" section (along with, optionally, a section for ethical considerations) may be up to one page and will not count toward the final page

limit. Note that these files may be used by venues that do not rely on ARR so it is recommended to verify the requirement of a "Limitations" section and other criteria with the venue in question.

## Acknowledgments

This document has been adapted by Steven Bethard, Ryan Cotterell and Rui Yan from the instructions for earlier ACL and NAACL proceedings, including those for ACL 2019 by Douwe Kiela and Ivan Vulić, NAACL 2019 by Stephanie Lukin and Alla Roskovskaya, ACL 2018 by Shay Cohen, Kevin Gimpel, and Wei Lu, NAACL 2018 by Margaret Mitchell and Stephanie Lukin, BibTeX suggestions for (NA)ACL 2017/2018 from Jason Eisner, ACL 2017 by Dan Gildea and Min-Yen Kan, NAACL 2017 by Margaret Mitchell, ACL 2012 by Maggie Li and Michael White, ACL 2010 by Jing-Shin Chang and Philipp Koehn, ACL 2008 by Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui, ACL 2005 by Hwee Tou Ng and Kemal Oflazer, ACL 2002 by Eugene Charniak and Dekang Lin, and earlier ACL and EACL formats written by several people, including John Chen, Henry S. Thompson and Donald Walker. Additional elements were taken from the formatting instructions of the *International Joint Conference on Artificial Intelligence* and the *Conference on Computer Vision and Pattern Recognition*.

## References

- Forty-Two AI Lab. Seed42lab/ai-paper-crawl. <https://huggingface.co/datasets/Seed42Lab/AI-paper-crawl>.
- Zhaowei Li, Qi Xu, Dong Zhang, Hang Song, YiQing Cai, Qi Qi, Ran Zhou, Juntong Pan, Zefeng Li, Vu Tu, Zhida Huang, and Tao Wang. 2024. [GroundingGPT: Language enhanced multi-modal grounding model](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6657–6678, Bangkok, Thailand. Association for Computational Linguistics.
- Yukiko Nakano, Gabe Reinstein, Tom Stocky, and Justine Cassell. 2003. [Towards a model of face-to-face grounding](#). In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, pages 553–561, Sapporo, Japan. Association for Computational Linguistics.
- Linhui Xiao, Xiaoshan Yang, Xiangyuan Lan, Yaowei Wang, and Changsheng Xu. 2024. [Towards visual grounding: A survey](#). *Preprint*, arXiv:2412.20206.

## A Example Appendix

This is an appendix.