Annotating APPRAISAL in text and corpora

Challenges and strategies

Matteo Fuoli

LUND UNIVERSITY

Mini-workshop on Appraisal Theory and the Language of Evaluation
Universitat Pompeu Fabra, 18th of March, 2016

Outline

1. Introduction: Evaluation in discourse

2. An overview of APPRAISAL theory

3. Challenges in annotating APPRAISAL in text

4. Strategies for reliably annotating APPRAISAL

Annotating APPRAISAL in text and corpora: Challenges and strategies, Matteo Fuoli, Lund University, Sweden

Introduction

Evaluation in discourse

- That was a **really great** movie!

- Donald Trump is a **clown**.

- **Sadly**, what we've all been hearing is true.

Bank of America

IKEA

Kellogg's

Nestlé

Volkswagen

bp

Kraft

NOVARTIS

Introduction
Evaluation in discourse

- BP is **progressive**, **responsible**, **innovative** and **performance driven**. [BP 2009 Sustainability Review]

- Our business idea is to offer a wide range of **well designed**, **functional** home furnishing products. [IKEA 2009 Sustainability Report]

Introduction

Evaluation in discourse

- Our business idea is to offer a wide range of **well designed**, **functional** home furnishing products.

Annotating APPRAISAL *in text and corpora: Challenges and strategies*, Matteo Fuoli, Lund University, Sweden

Annotating APPRAISAL *in text and corpora: Challenges and strategies*, Matteo Fuoli, Lund University, Sweden

An overview of APPRAISAL theory

Annotating APPRAISAL in text and corpora: Challenges and strategies, Matteo Fuoli, Lund University, Sweden

Background
Theoretical approaches to evaluation

- Stance (Biber and Finegan, 1988, 1989; Conrad and Biber, 2000)

- Evaluation (Hunston and Thompson, 2000; Thompson and Alba-Juez, 2014)

- APPRAISAL (Systemic Functional Linguistics) (Martin and White, 2005)

- Bednarek's parametric approach to evaluation (Bednarek, 2006)

- Stance-taking (Conversation Analysis) (Englebretson, 2007)

- Sentiment Analysis (Computational Linguistics) (Pang and Lee, 2008)

Annotating APPRAISAL in text and corpora: Challenges and strategies, Matteo Fuoli, Lund University, Sweden

Background

An overview of APPRAISAL theory

APPRAISAL

ATTITUDE → AFFECT, JUDGEMENT, APPRECIATION

ENGAGEMENT → MONOGLOSS, HETEROGLOSS

GRADUATION → FORCE, FOCUS
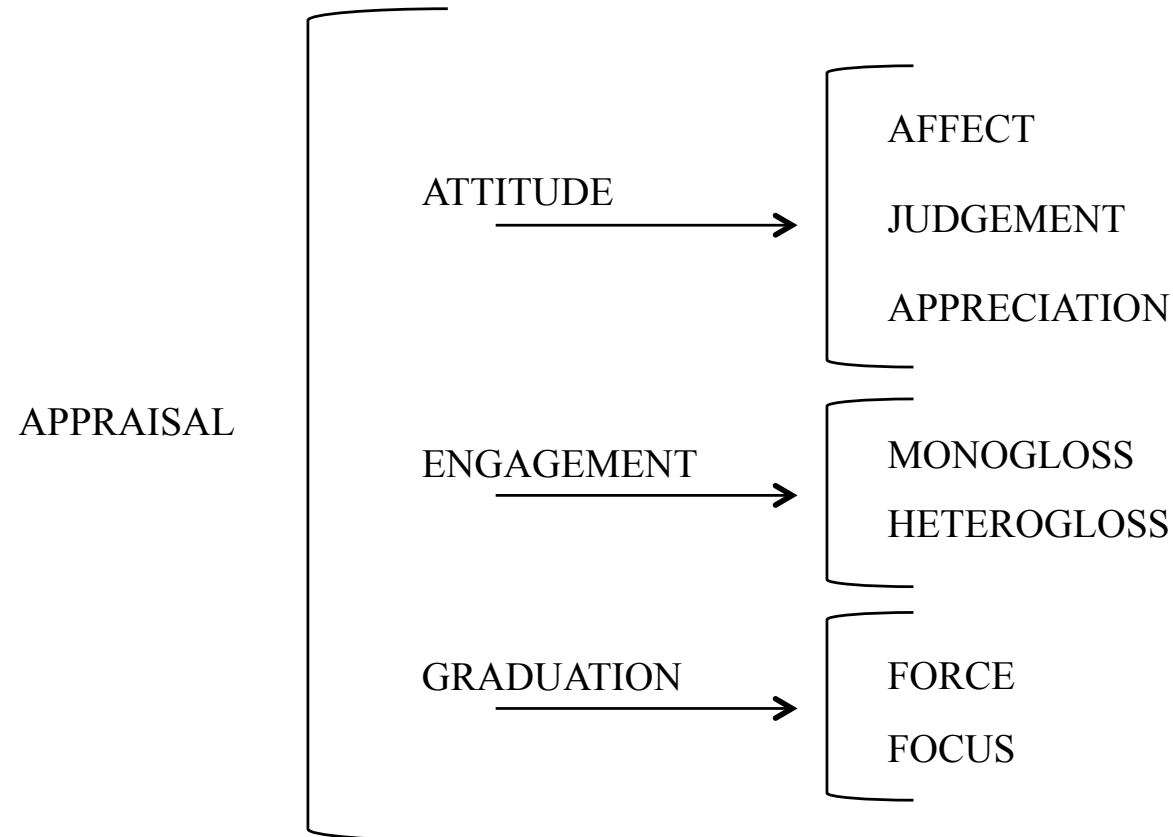
An overview of Appraisal (adapted from Martin and White, 2005)

Annotating APPRAISAL in text and corpora: Challenges and strategies, Matteo Fuoli, Lund University, Sweden

Background

An overview of APPRAISAL theory

- ATTITUDE
  - The explosion and fire on the Deepwater Horizon rig **shocked** everyone within BP and we feel great **sadness** that 11 people died.

- ENGAGEMENT
  - Despite these actions, ConocoPhillips **considers it possible** that the recession **could** restrain energy demand and prices for several years.

- GRADUATION
  - I am **extremely** proud of BP's 2009 safety performance.

Annotating APPRAISAL in text and corpora: Challenges and strategies, Matteo Fuoli, Lund University, Sweden

Background
Manual text annotation

- Manual text annotation has become a popular method for examining and comparing the use of evaluative language resources across texts and corpora (e.g. Bednarek, 2008; Carretero and Taboada, 2014; Fuoli, 2012; Fuoli and Hommerberg, 2015; Lipovsky, 2013; O'Donnell, 2014; Pounds, 2011; Taboada and Carretero, 2012; Ryshina-Pankova, 2014; Santamaría-García, 2014)

- However, annotating APPRAISAL presents several challenges to achieving reliable and replicable analyses

Challenges in annotating APPRAISAL

Annotating APPRAISAL *in text and corpora: Challenges and strategies*, Matteo Fuoli, Lund University, Sweden

Challenges in annotating APPRAISAL

- Evaluation may be realized through an open-ended range of expressions of varying length and complexity and belonging to any word class (e.g. Hunston, 2011)

> excels [VB]
> outstanding [ADJ]
> extremely talented [ADJP]
> in an enviably strong position [ADV]
> one of the world's great enterprises [Complex NP]

Challenges in annotating APPRAISAL

- Evaluation is a highly context-dependent phenomenon (e.g. Bednarek 2006: 8; Hunston 2011: 10; Macken- Horarik and Isaac 2014; Martin and White 2005; Paradis et al. 2012; Thompson and Alba-Juez 2014)

There's thin and light. Then there's **thin** and **light** on a whole new level. iPod touch has a super-thin aluminum body that feels barely there in your hand or pocket.

Challenges in annotating APPRAISAL

- Setting the textual boundaries of evaluative expressions (*unitizing*) is often a complicated and arbitrary exercise

We are well-positioned to generate shareholder value with distinct competitive advantages and a **steadfast commitment** to the **highest standards of ethics**.

Annotating APPRAISAL *in text and corpora: Challenges and strategies*, Matteo Fuoli, Lund University, Sweden

Challenges in annotating APPRAISAL

- Setting the textual boundaries of evaluative expressions (*unitizing*) is often a complicated and arbitrary exercise

> We are well-positioned to generate shareholder value with distinct competitive advantages and a **steadfast commitment to the highest standards of ethics**.

Challenges in annotating APPRAISAL

- The distinction between *inscribed* and *invoked* APPRAISAL is far from clear-cut (e.g. Bednarek, 2006)

> There's thin and light. Then there's **thin** and **light** on a whole new level. iPod touch has a super-thin aluminum body that feels barely there in your hand or pocket.

Challenges in annotating APPRAISAL

- Different interpretations for an expression are often equally plausible, and multiple category labels valid

> We are a self-sustaining and competitive international, integrated energy company with **diligent** financial management.

Annotating APPRAISAL in text and corpora: Challenges and strategies, Matteo Fuoli, Lund University, Sweden

Challenges in annotating APPRAISAL

- The distinction between the categories in the framework is not always clear

  ExxonMobil Chemical has delivered **industry-leading** <u>performance</u>.

Annotating APPRAISAL *in text and corpora: Challenges and strategies*, Matteo Fuoli, Lund University, Sweden

Challenges in annotating APPRAISAL

- Annotating APPRAISAL entails a number of practical issues

> ExxonMobil's success is underpinned by our commitment to integrity - our **systematic** and **unwavering** focus on safety.

Annotating APPRAISAL *in text and corpora: Challenges and strategies*, Matteo Fuoli, Lund University, Sweden

Challenges in annotating APPRAISAL

- Annotating APPRAISAL entails a number of practical issues

> ExxonMobil's success is underpinned by our commitment to integrity - our **systematic and unwavering** focus on safety.

Reliability, replicability, transparency

- **Reliability**: "the extent to which a measurement procedure yields the same answer however and whenever it is carried out" (Kirk and Miller, 1986: 19)
  - *test-retest* reliability
  - *internal consistency* reliability
  - *interrater* reliability

Annotating APPRAISAL in text and corpora: Challenges and strategies, Matteo Fuoli, Lund University, Sweden

Reliability, replicability, transparency

- The question of reliability has been dismissed by Martin and White on the grounds that "so-called objectivity is impossible" (Martin and White, 2005: 207)

- Measures of inter-coder or intra-coder agreement are rarely reported (for exceptions, see Fuoli, 2012; Fuoli and Hommerberg, 2015; Read and Carroll, 2010; Ryshina-Pankova, 2014; Taboada and Carretero, 2012)

- In certain cases limited information is given about the annotation criteria or about how this process was carried out (e.g. Lipovsky, 2008, 2013; Mackay and Parkinson, 2009; Pounds, 2010, 2011; Santamaría-García 2014)

A step-wise method for annotating APPRAISAL

Annotating APPRAISAL in text and corpora: Challenges and strategies, Matteo Fuoli, Lund University, Sweden

1. Define the scope of the project

⇩

2. Select and configure an annotation tool

⇩

3. Draft a context-specific annotation manual

⇩

4. Assess reliability

⇩

5. Refine the annotation manual

⇩

6. Annotate the corpus

⇩

7. Analyze the results

Annotating APPRAISAL *in text and corpora: Challenges and strategies*, Matteo Fuoli, Lund University, Sweden
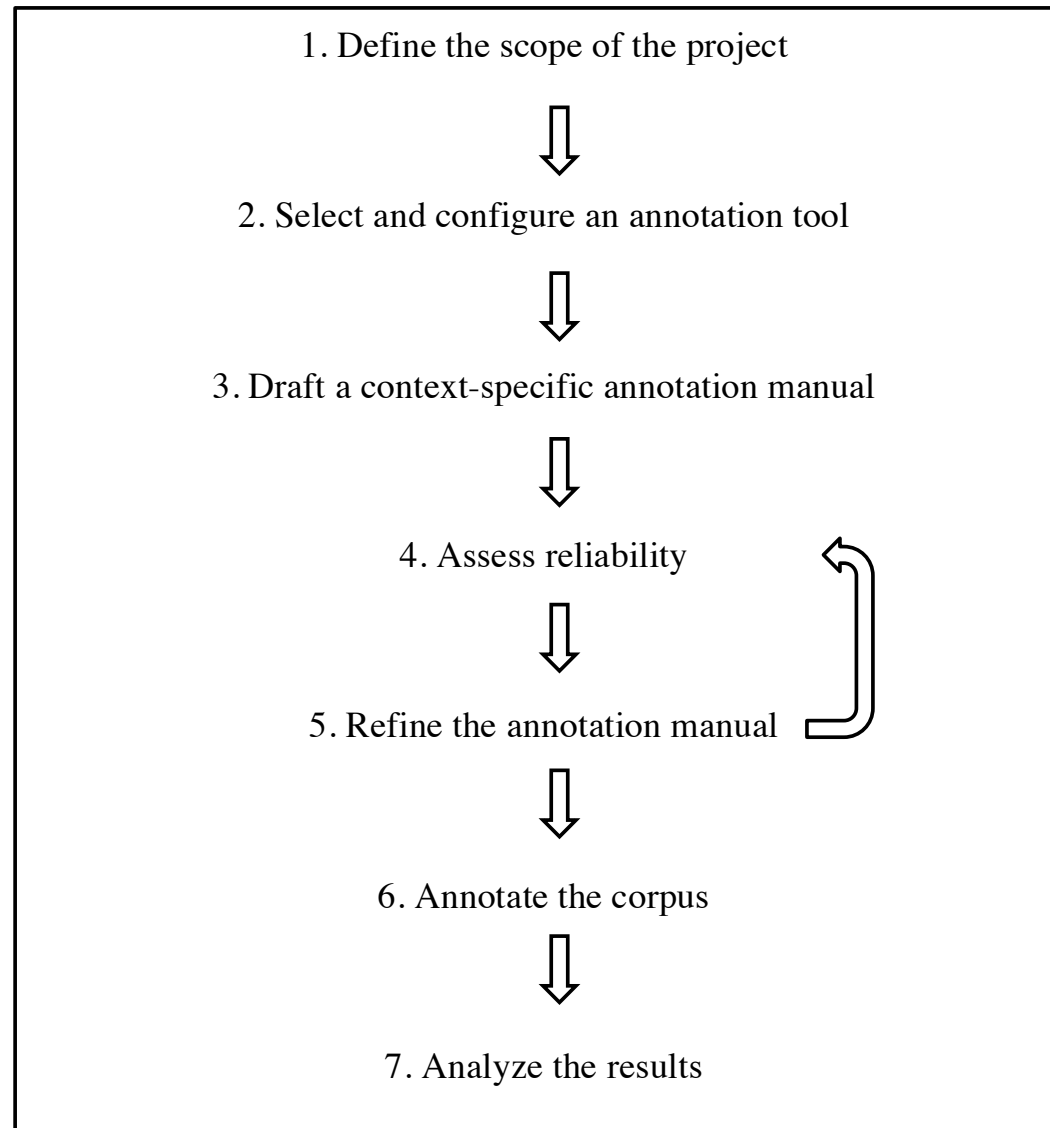
General principles

- PRINCIPLE 1. All choices should be accounted for

- PRINCIPLE 2. The annotation guidelines should be tested and refined until maximum reliability is achieved

- PRINCIPLE 3. Reliability should always be assessed, and reliability scores reported and discussed

Step 1:
Define the scope of the project

- Create a preliminary version of the coding scheme
  - categories should be chosen based on the research question(s)
  - not all APPRAISAL categories necessarily need to be included
- Informally annotate random samples from the corpus
  - take note of any unexpectedly frequent or interesting phenomena
- Add or remove categories from the draft coding scheme

⬇
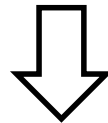
Preliminary list of the APPRAISAL categories and subcategories to be included in the annotation scheme

Step 2:
Choose and configure an annotation tool

- Several freeware programs are available (for a review, see O'Donnell 2014)

  - UAM CorpusTool (O'Donnell, 2008)

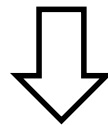 - CAT annotation tool (Lenzi et al., 2012)

Step 3:
Draft a *context-specific* annotation manual

- The manual should (minimally) include

  - outline of the annotation scheme

  - context-specific category definitions

  - explicit rules for applying the definitions to the data set under study

  - illustrative examples

⬇

The manual should be **context-specific**: the definitions and guidelines should be shaped around the specific characteristics of the texts to be annotated

Step 4:
Assess reliability

- To assess *test-retest reliability*:
  - annotate a random sample from the corpus twice, allowing some time between the two sessions
  - compare the annotations and calculate *intra*-coder agreement

- To assess *interrater reliability*:
  - annotate a random sample from the corpus
  - ask one or more other people to annotate the same data independently, following the guidelines specified in the annotation manual
  - compare the annotations and calculate *inter*-coder agreement
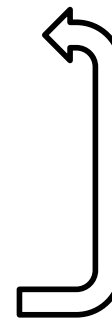
Step 5:
Refine the annotation manual

- The results of the tests should be used to improve and refine the annotation guidelines as to optimize their robustness and clarity

- Reliability should be tested again, and the guidelines refined, until a satisfactory and/or stable level of agreement has been reached

4. Assess reliability

5. Refine the annotation manual

Step 6:
Annotate the corpus

- This task can be performed by one or more people, provided that interrater reliability is high

- Steps should be taken to minimize fatigue and cognitive load, which may negatively affect consistency
  - reviewing each text multiple times can contribute to achieving higher internal consistency

Step 7:
Analyze the results

- Once the corpus has been fully annotated, the data can be analyzed using both quantitative and qualitative techniques

- Quantitative analysis may be carried by simply counting and comparing the frequency of APPRAISAL categories across texts, but also using more sophisticated multifactorial techniques
  (see, e.g. Divjak, 2006; Geeraerts et al., 1994; Glynn, 2009; Gries, 1999)

Conclusion

- Practical solutions to the problem of subjectivity, which may hinder the reliability and replicability of analyses and have a major impact on the results

- By creating explicit and detailed annotation guidelines and refining them until optimal reliability is reached, the subjectivity involved in annotating APPRAISAL is not eliminated, but 'tamed' and controlled for, to the maximum extent possible

Limitations

- It is comparatively more time consuming than a more informal approach to manual annotation

- It may be better suited to quantitative than qualitative analysis
  - more coarse-grained and rigid than the qualitative, interpretive approach adopted by many scholars working with the APPRAISAL framework (e.g. Macken-Horarik and Isaac, 2014)
  - complementary approaches?

- General criteria and standard thresholds for reliability have not been established yet for the task of annotating APPRAISAL, so it is unclear when an analysis can be considered reliable *enough*

- Fuoli, M. (*in press*). A step-wise method for annotating APPRAISAL. *Functions of Language*, 25(1). [Manuscript available via Academia.edu]

- Fuoli, M. and Hommerberg, C. (2015). Optimizing transparency, reliability and replicability: Annotation principles and inter-coder agreement in the quantification of evaluative expressions. *Corpora*, 10(3).

- Slides from my talk available via Academia.edu [including complete references]

Thanks for listening

Matteo Fuoli
Lund University
matteo.fuoli@englund.lu.se

References [1]

- Bednarek, M. 2006. Evaluation in Media Discourse: Analysis of a Newspaper Corpus. London and New York: Continuum International Publishing Group Ltd.
- Bednarek, M. 2008. Emotion Talk across Corpora. Houndmills, Basingstoke: Palgrave Macmillan.
- Biber, D. and E. Finegan. 1988. 'Adverbial stance types in English', Discourse Processes 11 (1), pp. 1–34.
- Biber, D. and E. Finegan. 1989. 'Styles of stance in English: lexical and grammatical marking of evidentiality and affect', Text- Interdisciplinary Journal for the Study of Discourse 9 (1), pp. 93–124.
- Carretero, M. and M. Taboada. 2014. 'Graduation within the scope of attitude in English and Spanish consumer reviews of books and movies' in G. Thompson and L. Alba-Juez (eds) Evaluation in Context, pp. 221–39. Amsterdam and Philadelphia: John Benjamins.
- Conrad, S. and D. Biber. 2000. 'Adverbial marking of stance in speech and writing' in S. Hunston and G. Thompson (eds) Evaluation in Text: Authorial Stance and the Construction of Discourse, pp. 56–73. Oxford: Oxford University Press.
- Divjak, D. 2006. 'Ways of intending: Delineating and structuring near synonyms' in S. T. Gries & A. Stefanowitsch (eds) Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis, pp. 19–56. Berlin: Mouton.
- Englebretson, R. 2007. Stancetaking in discourse: subjectivity, evaluation, interaction. Amsterdam and Philadelphia: John Benjamins.
- Fuoli, M. 2012. 'Assessing social responsibility: a quantitative analysis of Appraisal in BP's and Ikea's social reports', Discourse and Communication 6 (1), pp. 55–81.
- Fuoli, M. and C. Hommerberg. 2015. 'Optimizing transparency, reliability and replicability: Annotation principles and inter-coder agreement in the quantification of evaluative expressions', Corpora, 10(3), pp. 315–349.

References [2]

- Geeraerts, D., S. Grondelaers and P. Bakema. 1994. The structure of lexical variation. Berlin: Mouton.
- Glynn, D. 2009. 'Polysemy, syntax, and variation. A usage-based method for Cognitive Semantics' in V. Evans and S. Pourcel (eds) New Directions in Cognitive Linguistics, pp. 77–104. Amsterdam: Benjamins.
- Gries, S. T. 1999. 'Particle movement: A cognitive and functional approach', Cognitive Linguistics, 10(2), pp. 105–146.
- Hunston, S. 2011. Corpus Approaches to Evaluation: Phraseology and Evaluative Language. New York and London: Routledge.
- Hunston, S. and G. Thompson. 2000. Evaluation in Text: Authorial Stance and the Construction of Discourse. Oxford: Oxford University Press.
- Kirk, J. and M. L. Miller. 1986. Reliability and validity in qualitative research. Beverly Hills, CA: Sage.
- Lipovsky, C. 2008. 'Constructing affiliation and solidarity in job interviews', Discourse & Communication, 2(4), pp. 411-432.
- Lipovsky, C. 2013. 'Negotiating one's expertise through appraisal in CVs', Linguistics and the Human Sciences 8 (3), pp. 307–33.
- Martin, J. and P. White. 2005. The Language of Evaluation: Appraisal in English. London and New York: Palgrave Macmillan.
- Mackay, J. and J. Parkinson. 2009. '"My very own mission impossible": An APPRAISAL analysis of student teacher reflections on a design and technology project', Text & Talk, 29(6), pp. 729–753.

Annotating APPRAISAL in text and corpora: Challenges and strategies, Matteo Fuoli, Lund University, Sweden

References [3]

- Macken-Horarik, M. and A. Isaac. 2014. 'Appraising Appraisal' in G. Thompson and L. Alba-Juez (eds), Evaluation in context, pp. 67–92. Amsterdam: Benjamins.
- O'Donnell, M. 2014. 'Exploring identity through appraisal analysis: a corpus annotation methodology', Linguistics and the Human Sciences 9 (1), pp. 95–116.
- Pang, Bo & Lillian Lee. 2008. 'Opinion mining and sentiment analysis', Foundations and Trends in Information Retrieval 2(1–2).
- Paradis, C., J. van de Weijer, C. Willners & M. Lindgren. 2012. 'Evaluative polarity of antonyms', Lingue e linguaggio, 11(2), pp. 199–214.
- Pounds, G. 2010. 'Attitude and subjectivity in Italian and British hardnews reporting: The construction of a culture-specific 'reporter' voice', Discourse Studies, 12(1), pp. 106–137.
- Pounds, G. 2011. "'This property offers much character and charm": evaluation in the discourse of online property advertising', Text and Talk 31 (2), pp. 195–220.
- Read, J. and J. Carroll. 2012. 'Annotating expressions of Appraisal in English', Language Resources and Evaluation, 46(3), pp. 421–447. First published online, December 2010.
- Ryshina-Pankova, M. 2014. 'Exploring argumentation in course-related blogs through engagement' in G. Thompson and L. Alba-Juez (eds) Evaluation in Context, pp. 281–302. Amsterdam and Philadelphia: John Benjamins.
- Santamaría-García, C. 2014. 'Evaluative discourse and politeness in university students' communication through social networking sites' in G. Thompson and L. Alba-Juez (eds) Evaluation in Context, pp. 387–411. Amsterdam and Philadelphia: John Benjamins.
- Taboada, M. and M. Carretero. 2012 [2010]. 'Contrastive analyses of evaluation in text: key issues in the design of an annotation system for attitude applicable to consumer reviews in English and Spanish', Linguistics and the Human Sciences 6 (1–3), pp. 275–95.
- Thompson, G. and L. Alba-Juez. 2014. Evaluation in Context. Amsterdam and Philadelphia: John Benjamins.