Merging Community Knowledge and Self-Interest to Build Language Resources: Architecture and Quality Management of a 'Take-and-Share' Approach of Linguistic Annotations

Hagen Peukert[1]

ZFDM, Universität Hamburg[1]

Abstract

Linguistic annotations of texts are time- and resource-intensive and hence expensive. Even if sufficient funds are supplied for manual tagging – still the gold standard of annotating texts – it remains an error-prone process, in which quality control soon reaches its limits. In addition, often a very limited number of users are in need of particular annotations required for very particular research questions so that economies of scale and scope of a larger research community could not easily be exploited. This paper addresses this issue taking into account research from Social Psychology considering the specific properties of texts. As a result of the interdisciplinary analyses, the design of a web architecture is suggested that has the potential of overcoming the above mentioned dilemma and significantly improve the quality of linguistic text annotations.

Veröffentlicht durch die Gesellschaft für Informatik e.V. 2018 in.

M. Burghardt & C. Müller-Birn (Hrsg.): INF-DH 2018 – Workshopband, 25. Sept. 2018, Berlin https://doi.org/10.18420/infdh2018-01

1 Introduction

More recently, essentials of human behavior have been considered to a reasonable extent in designing webpages and software agents leading to what is nowadays called human-centered design (Butz and Krüger, 2014). Usage behavior and consumer behavior have particularly played an important role from the very early days of software and web technologies to drive up profits in new market segments of online services and selling (Massit-Folléa and Méadel, 2006; Sandholm, 2007). These approaches have undergone further development under the influence of research carried out in Artificial Intelligence (N. J. Nilsson, 2009; Russell and Norvig, 2010; Poole and Mackworth, 2010).

Thus the technological enhancement moves the human being per se to the center stage of software engineering and aligns consistently the functionality of software components to the needs of human behavior. This also means that the underlying assumption of rational choice widely present in early models falls behind the logics of human personality (Russell, 1997). As a consequence more knowledge from the psychological sciences is entering the field of software engineering and development. It is a trend that must be most welcome in disciplines on the intersection between technology and human behavior. It comes as no surprise that the investigation of language as an exclusively human property and one of the best reflectors of both, traits and culture of human behavior, as well as research in language technologies are particularly supportive in this direction.

If further advances bring an even more personality-centered approach to perfection (O. Nilsson, 2006), it is about time for language technologies to be more consistently equipped with theories of Differential Psychology, which studies human personality traits, but also Social Psychology, which investigates human behavior within the societal setting, e.g. *ingroup* behavior. Knowledge of these domains should enable scientist of language technology to design more elaborated automated language processing systems viewing human behavior as it is as opposed to how it should be.

So, the tendency described above also applies to software that helps to create (language) resources. Even more so, generating language resources usually involves, beside a lot of manpower, processes that cannot easily be automated and a different set of requirements of staff motivation and organization. Needless to say that knowledge on human behavior plays an important role when developing tools that lend the user a hand in annotating texts. In short, this knowledge should be directly or indirectly used in devising the software.

In the suggested architecture below, a human-centered view is taken in the conceptual design of a community platform for deep annotations of larger text corpora. On the one hand, the main idea of the architecture focuses on the general repugnance of humans to grant a free lunch even at the cost of an individual disadvantage (Fehr and Gächter, 2002; Gintis, 2000). In addition, humans have an inherent interest in community and social well-being and they are generally willing to behave altruistically under strict conditions (Omoto and Snyder, 2010). On the other hand, the model at hand makes use of the specific properties of texts and describes how these properties could be exploited in overcoming obstacles in quality management. Without these properties, that is a diminishing marginal vocabulary growth of texts in general (Baayen, 1996),

the architecture could not work. For that reason, it makes little sense to convey the principle to the processing of other data formats. It is restricted to language processing and thus language technology.

2 The Problem of Quality Management in Building Annotations: Human Social Behavior and Personality Traits

The approach suggested here is reminiscent of web wikis, i.e. a critical mass of users contribute bits and in this way all users help to build up, by and large equally, a resource of substantial size, which makes it attractive to the entire community, but also to the individual despite the costs of investment. Yet, set aside ownership guarantees and some problems with terms and conditions, there are two fundamental challenges to the community approach on the strategic level nowadays: how to achieve the critical mass of active users and how to assure quality of their work.

First, a short look at the many deserted less prominent web community projects reveals that, in absence of force, other factors but mere altruism are at play and drive community projects to success or loss. It is, thus, essential to know the major impacting factors and manipulate them accordingly. Leaving some minor exceptions beside, in the majority of all community projects, it is a concrete advantage that the user needs to see before an involvement in the project takes place (Massit-Folléa and Méadel, 2006). These advantages are manifold and cannot be discussed in depth as deserved. Suffice it to say that individual advantages may not be salient although they might often appear so as predicted by social and group identity theory (Postmes and Spears, 1998; Gaertner and Dovidio, 2000; Tajfel and Turner, 2004). To be precise, writing wikipedia articles improves the author's reputation and reference, helps spreading particularly her or his own ideas as a inexpensive way of publication or satisfies the need for self-esteem. Notwithstanding most of the time, the author does not know about these real motives and believes it is mere empathy to help others getting to know a topic. In sum, achieving involvement of a sufficient number of dedicated users depends on the interaction of empathy, self- and cooperative interests (Stürmer and Snyder, 2010).

Second, quality control is an issue and well-known from endless projects in which manually tagged corpora had to be prepared. More specifically, creating linguistic annotations of texts reveals a Janus-faced task. On the one side, deep annotations call for a high degree of expertise, which is to be found at rare white-collar segments disposing of high qualification and specific perceptions on the job tasks and self-fulfillment. Conceptions of salary levels are little negotiable and, in fact, a bad motivator in the long run anyways (Nerdinger, 2014). On the other side, annotation work in general and word annotations in particular are tedious and extraordinarily tiring especially if carried out for a longer time such as weeks or months (Omoto and Snyder, 2010; Grube and Piliavin, 2000). Research on the initial motivation of newcomers in a community converges with the established effects (Balestra et al., 2017). Still, very conscientious work is indispensable.

These janiform job requirements – coping with a monotonously dreary task and expecting highly qualified personnel at a given salary level – run counter to the expectancy function of potential employees (Van den Broeck et al., 2010). Put briefly, it is a classic problem of missing positive incentives (Lee et al., 1999). It follows that suboptimal compromises have to be accepted in presence of plentiful job alternatives and given that in the majority of research projects management concepts such as job enrichment and job rotation cannot be put into practice due to limited staff and flat hierarchies in the organizational structure (K. Fuchs-Kittowski and F. Fuchs-Kittowski, 2002; Sonnentag and Frese, 2002).

Moreover, control scenarios aiming at a direct surveillance or re-checking results even if based on samples fail for the same reason. In fact, the perception of employee control scenarios impacts stronger on demotivation than an monotonous task itself (Nerdinger, 2014). However, employees tend to cheat to get over a reluctant task unless staff conceives of the task as added value attributable directly to oneself and recognized as such by others. If these two conditions are met employees are willing to invest about the same time and workload as if working for their own advantage only (Lundin and Rasmussen, 2002; Schaufeli and Bakker, 2010).

3 Properties of Text as a Gatekeeper for a *Take-and-Share* Approach

To understand why the present approach has good chances to work, it is necessary to take a look at the distribution and occurrence of words in large texts. There are two qualities of texts that are important in the present context. First, measuring the frequency of new word types dependent on text size across all registers shows a clear picture. Known as Zipf's law (Zipf, 1929), the type frequency diminishes with a growing text size, that is, the function of type frequency takes a concave form, i.e. the TTR (type-token ratio) deacreases (Baayen, 1996; Baayen and Renouf, 2001; Baayen, 2001). This also means that if a critical mass of word types is reached, the probability of encountering a new word decreases with each additional unit of text. Second, content word types of a general low occurrence in texts show up in clusters (Stubbs, 2001, p. 310).

Thus a text can be subdivided into words that are peculiar to this text, i.e. the probability of occurrence of these types is higher than in any other text and the rest of words whose probability of occurrence is about equal in all texts. The share of the peculiar words of a text is equal to the amount of annotation work to be carried out by the user. With each additional annotated text-specific word added to a stock of data that is used to automatically map subsequent words, potential future annotation work decreases. However, it has been shown that the number of text-specific words remains relatively robust over different texts (Baayen, 2001).

In analogy to the wikipedia approach, the approach at hand suggests that an author would only need to write about a paragraph and could announce a full-fledged article. By the same ratio, the annotator only works on a small fraction of the entire text. So the first appeal is a diminished workload of considerable size. The second appeal is, at least up to now, the absence of good alternatives. There is no cheaper way to receive word annotations of this quality. However, the

present approach differs from the wikipedia analogy in that no one can access the contents of annotations without ever actively contributing to its overall stock.

4 Multistage Design as a Collateral for Quality Assurance: Converging Text Properties and Human Behavior

The essence of the *take-and-share* approach introduced here involves two stages: performance and reward. The first stage comprises work in advance in the form of manual annotations and the second stage ensures the reward, which is a fully annotated text. The software, whose architecture is to be outline here, grants access to the fully annotated text only after all annotation work is completed. The amount of annotation work is indeed hardly ever exactly the same for all users because text length and the number of hapax legomena, neologisms, and words not listed in the data base will differ – predicted by Zipf's law – as residuals disperse around a true theoretical value. Since the software assures that everyone, using the mapping of word annotation to one's own text, must contribute about equally and at the same time contributes to the stock of linguistic annotation, from which the entire community will profit, no social loafing (Latané et al., 1979) is permitted at no time to anyone. As a side effect, other motivation losses as the *dispensability effect* Kerr and Bruun, 1983 and the *sucker effect* (Kerr, 1983) are also avoided and motivation gains are automatically promoted (*social competition* and *Köhler effect* (Stroebe et al., 1996; Witte, 1989). In how far *social compensation* (Williams and Karau, 1991) is significant cannot be evaluated at present.

Motivation is held up because each user is directly affected by the quality of the analysis and only a relatively small share of the work has to be done compared to the alternative, that is, annotating the entire text, without previous mapping from the master database. It is clear that the workload of annotating a relatively small corpus of a $100,000$ words or so is immense for a single researcher and therefore it is the better choice given that no other individual profits more than oneselve from the provided annotation work (Fehr and Gächter, 2002; Gintis, 2000). One can attribute the effect of the diminished workload directly to the properties of text to the diseconomies of scale of word types (Baayen, 1996) because it means that the mass of text length is due to word tokens whose annotations are already known and which can easily be mapped to an input text. Taken together with the cluster effect of low-frequency word types (Stubbs, 2001) the workload diminishes even further since the overall text size – if measured on the basis of word types – does reduce substantially.

In other words, a community approach such as a *take-and-share* model is especially beneficial for annotations of texts. While a non-community approach would face a rather disproportionate increase of annotation work for the very reason of taking care of very specific text genres and vocabularies, a community approach makes use of leaving all idiosyncrasies unconsidered. These cases that are specific to single texts can be outsourced to the respective user who is still profiting from saving time and work for having the mass of word types automatically mapped to a repository built by the user community.

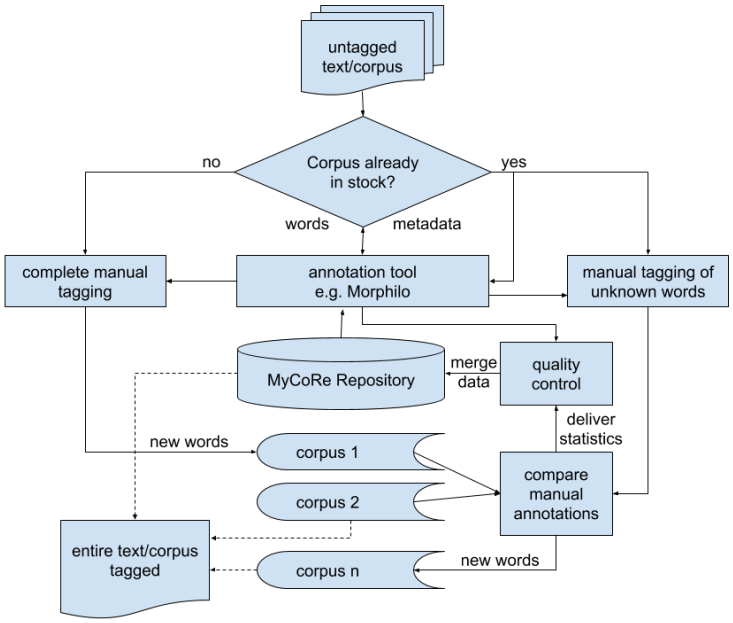

Figure 1: Basic Architecture of a 'Take-&-Share' Approach

5 Details of the Web-Architecture

The architecture of a possible *take-and-share* approach for language resources is visualized in figure 1. Because the very gist of the approach becomes clearer if describing a concrete example, the case of annotating lexical derivatives of Middle English with the help of the Morphilo Tool (Peukert, 2012) using a *MyCoRe* repository[1] is given as an illustration. However, any other tool that helps with manual annotations and manages metadata of a corpus could be substituted here instead.[2]

After inputting an untagged corpus or plain text, it is determined whether the input material was annotated previously by a different user. This information is usually provided by the metadata administered by the annotation tool; in the case at hand, the *Morphilo* component. An alternative is a simple table look-up for all occurring words in the datasets Corpus 1 through Corpus n. If contained completely, the *yes*-branch is followed up further – otherwise *no* succeeds. The difference between the two branches is subtle, yet crucial. On both branches, the annotation tool (here *Morphilo*) is called, which, first, sorts out all words that are not contained in the master database (here *MyCoRe* repository) and, second, makes reasonable suggestions on an optimal annotation of the items. The suggestions made to the user are based on simple

[1] http://www.mycore.de

[2] The source code of a possible implementation is available on https://github.com/amadeusgwin/morphilo. The software runs in test mode on https://www.morphilo.uni-hamburg.de/content/index.xml, however a login is required that can be granted by the author.

string mapping of a saved list of prefixes and suffixes whereas the remainder of the mapping is defined as the word root. The annotations are linked to the respective items (e.g. words) in the text, but they are also persistently saved in an extra dataset, i.e. in figure 1 in one of the delineated Corpus 1 through n, together with all available metadata.

The difference between the two branches in figure 1 is that in the *yes*-branch a comparison between the newly created dataset and all of the previous datasets of this text is carried out while this is not possible if a text was not annotated before. Within this unit, all deviations and congruencies of the annotated items are marked and counted. The underlying assumption is that with a growing number of comparable texts the correct annotations approach a theoretic true value of a correct annotation while errors level out provided that the sample size is large enough. How the distribution of errors and correct annotations exactly looks like and if a normal distribution can be assumed is still object of the ongoing research, but independent of the concrete results, the component (called *compare manual annotations* in figure 1) allows for specifying the exact form of the sample population. In fact, it is necessary at that point to define the form of the distribution, sample size, and the rejection region. To be put it simple here, a uniform distribution in form of a threshold value of e.g. 20 could be defined that specifies that a word has to be annotated equally by 20 different users before it enters the master database.

Continuing the information flow in figure 1 further, the threshold values or, if so defined, the results of the statistical calculation of other distributions respectively are delivered to the quality-control-component. Based on the statistics, the respective items together with the metadata, frequencies, and, of course, annotations are written to the master database. All information in the master database is directly used for automated annotations. Thus it is directly matched to the input texts or corpora respectively through the *Morphilo*-tool. The annotation tool decides on the entries looked up in the master which items are to be manually annotated.

The processes just described are all hidden from the user who has no possibility to impact the set quality standards but by errors in the annotation process. The user will only see the number of items of the input text he or she will process manually. The annotator will also see an estimation of the workload beforehand. On this number, a decision can be made if to start the annotation at all. It will be possible to interrupt the annotation work and save progress on the server. And the user will have access to the annotations made in the respective dataset, correct them or save them and resume later. It is important to note that the user will receive the tagged document only after all items are fully annotated. No partially tagged text can be output.

6 Discussion

The most obvious disadvantage of the approach suggested above and a seemingly contradiction is that the initiator of a language resource project must provide the master repository containing a critical mass of annotated material. Depending on the kind of annotation this might comprise from several $1,000$ up to a $20,000$ items as a rough estimation. And these items have to meet highest quality standards so that only manual annotations come into consideration. Without question, the initiator now faces exactly the same problem as described above including all

disadvantages of work motivation and quality assurance. In fact, her or his investment are fixed costs of the project and there is no other possibility avoiding it.

A second problem when it comes to creating word annotations of Middle English is that usually a fairly small research community is involved although the number of projects demanding different kinds of deep annotations is large. This is just another dimension of the *long-tail*-problem of data in the humanities. Yet, it is relevant in the given context because the success of a *take-and-share* approach turns on the quantity of potential users. This means that before starting to build a master repository and possibly the implementation of an annotation tool, it is important to have a good idea on the number of people likely to use it. Since scientists normally have a concrete idea of the main players in their field of expertise, a realistic estimate should be made. The decision of implementing the suggested architecture crucially depends on the estimation of the number of potential users. Otherwise it is more promising to think about alternative models of distributed work sharing that are more appropriate in smaller research communities (de Vries, 2003). The above argument is not applicable, however, if a stock of data already exists, e.g. from previous projects, that is still incomplete or needs to be enlarged. For these cases, the proposal at hand, seems to be an optimal solution.

Quality assurance in *take-and-share* models emanate from the idea that a researcher involved in annotation work has a vital interest in the quality of the annotated text because the delivered quality directly feeds back into his or her own results and the trustworthiness of the research carried out. Based on this assumption, it is plausible to suppose an normally distributed error function, that is, misconceptions of an annotation happen randomly over the annotated items. Therefore, taken all participants' annotation work into account, wrong annotations cannot be related systematically to one item or participant but potentially to all occurring items in the population. This assumption is prerequisite to calculate the error as it is a basic assumption in inferential statistics. However, the argument could be raised that certain items such as a complex word is particularly tricky to tag. To be concrete, if a lexical annotation has to be attributed to "transmittance", the annotator might find that the complex could be split into prefix "trans" and the suffix "ance" leaving "mit" as a root. Since morphological theory is not clear on this issue, a systematic – and not random – error might occur. So one would expect some skewing in the data. Still, it is reasonable to expect to receive measurable deviations even in the cases of systematic skewing, i.e., that at least about 5 percent of the users will have deviant results so that at the given confidence level the skewed data is not to be carried over in the master database. Yet there is a small chance that it still does.

Considering the architecture in figure 1, it is a comprehensible proposal to grant more responsibility to the user in the form of giving the choice to select the datasets (Corpus 1 through n) on one's own risk before they are checked for quality (Peukert, 2014; Dowling and Nicholson, 2002). The advantage is that fewer annotation work has to be done because the potential stock of annotations is larger. Provided that admitting additional datasets is used with discernment, this option could give a further hint for quality control. The logic behind that approach is to count the most frequently selected datasets while assuming that usage frequency correlates with superior quality, that is, scientist are unlikely to select datasets from which they do not know how trustworthy they are. Albeit the idea seems attractive at first sight, one could as well

doubt the underlying assumption. It is a temptation to set additional information aside especially if they can save some time. This would mean that a critical mass of researchers is not as cautious as one would expect them to be. The decisive difference to the calculation of the error function in the above scenario is that in the latter case there is no hint at how the behavior of researchers can be assumed to be normally distributed or evaluated at all. At this point, it needs more experimental work with an implementation of the suggested architecture to know that. As long as no other cues are provided, it is best to be on the save side and disallow risking the use of questionable datasets. What could be done, however, is to use the datasets as suggestions in the Annotation Tool. This is a subtle difference because these annotations are still manually re-checked.

7 Summary

The paper suggests an architecture for annotating words based on a *take-and-share* approach of data processing. The architecture draws on the specific qualities of the distribution of word types in texts and so it is, to the best of the author's knowledge, only feasible with language data on the word level. Based on the text qualities, results from Social Psychology refine the architecture's underpinnings and motivate the idea that every user has to contribute in due proportion to his or her benefit first and only then receive an adequate reward – that is a fully annotated text. The advantage for the user is that despite a minimal investment of annotation work, a tremendous amount of time and work is saved. Moreover no other user has the chance to profit in undue proportion. Last, the quality of the annotation is maximized by two principles. First, researchers have an inherent interest in the correctness of their own work and, second, provided that the community is large enough, errors are leveled out by a large enough sample contributed by the community.

References

Baayen, H. (1996). The effects of lexical specialization on the growth curve of the vocabulary. *Computational Linguistics*, *22*, 455–480.

Baayen, H. (2001). *Word frequency distributions*. Dordrecht: Springer Science+Business Media.

Baayen, H. & Renouf, A. (2001). A corpus-based study of compounding in english. *Journal of English Linguistics*, *29*, 101–123.

Balestra, M., Cheshire, C., Arazy, O., & Nov, O. (2017). Investigating the motivational paths of peer production newcomers. In ACM (Ed.), *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 6381–6385). CHI. Denver.

Butz, A. & Krüger, A. (2014). *Mensch-maschine-interaktion*. München: Oldenbourg.

de Vries, S. (2003). Online knowledge communities: Meeting places for continuing professional development. In T. J. van Weert & R. K. Munro (Eds.), *Informatics and the digital society: Social, ethical and cognitive issues* (pp. 215–224). The International Federation for Information Processing. Dordrecht: Kluwer Academic Publishers.

Dowling, C. & Nicholson, P. (2002). Choice and responsibility: The delegation of decision making to intelligent software agents. In K. Brunnstein & J. Berleur (Eds.), *Human choice and computers: Issues of choice and quality of life in the information society* (pp. 163–170). The Internatinal Federation for Information Processing. Dordrecht: Kluwer Academic Publishers.

Fehr, E. & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Fuchs-Kittowski, K. & Fuchs-Kittowski, F. (2002). Quality of working life, knowledge-intensive work processes and creative learning organisations: Information processing paradigm versus self-organisation theory. In K. Brunnstein & J. Berleur (Eds.), *Human choice and computers: Issues of choice and quality of life in the information society* (pp. 265–274). The Internatinal Federation for Information Processing. Dordrecht: Kluwer Academic Publishers.

Gaertner, S. L. & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia: Psychological Press.

Gintis, H. (2000). Strong reciprocity and human sociality. *Journal of Theoretical Biology*, *206*, 169–179.

Grube, J. A. & Piliavin, J. A. (2000). Role identity, organizational experiences, and volunteer performance. *Personality and Social Psychology Bulletin*, *26*, 1108–1119.

Kerr, N. L. (1983). Motivation losses in small groups: A social dilemma analysis. *Journal of Personality and Social Psychology*, *45*, 819–828.

Kerr, N. L. & Bruun, S. E. (1983). Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and Social Psychology*, *44*, 78–94.

Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, *37*, 822–832.

Lee, L., Piliavin, J. A., & Call, V. R. A. (1999). Giving time, money, and blood: Similarities and differences. *Social Psychology Quarterly*, *62*, 276–290.

Lundin, M. S. & Rasmussen, L. B. (2002). A radical scandinavian ("øresundsk") approach to inquiring organizations: A critique of ict in knowledge management. In K. Brunnstein & J. Berleur (Eds.), *Human choice and computers: Issues of choice and quality of life in the information society* (pp. 293–303). The Internatinal Federation for Information Processing. Dordrecht: Kluwer Academic Publishers.

Massit-Folléa, F. & Méadel, C. (2006). Consumer models in the encounter between supply and demand of electronic administration. In J. Berleur, M. I. Nurminen, & J. Impagliazzo (Eds.), *Social informatics: An information society for all? in remembrance of rob kling* (pp. 209–217). International Federation for Information Processing. Oxford: Springer.

Nerdinger, F. W. (2014). Motivierung. In H. Schuler & U. P. Kanning (Eds.), *Lehrbuch der personalpsychologie* (pp. 725–764). Göttingen: Hofgrefe.

Nilsson, N. J. (2009). *The quest for artificial intelligence: A history of ideas and achievements*. Cambridge: Cambridge University Press.

Nilsson, O. (2006). A user centred access model. In J. Berleur, M. I. Nurminen, & J. Impagliazzo (Eds.), *Social informatics: An information society for all? in remembrance of rob kling* (pp. 445–455). International Federation for Information Processing. Oxford: Springer.

Omoto, A. M. & Snyder, M. (2010). Influences of psychological sense of community on voluntary helping and prosocial action. In S. Stürmer & M. Snyder (Eds.), *The psychology of prosocial behavior: Group prozesses, intergroup relations, and helping* (pp. 223–243). Chichester: Wiley-Blackwell.

Peukert, H. (2012). From semi-automatic to automatic affix extraction in middle english corpora: Building a sustainable database for analyzing derivational morphology over time. In J. Jancsary (Ed.), *Empirical methods in natural language processing* (pp. 413–423). Scientific Series of the ÖGAI. Wien.

Peukert, H. (2014). The morphilo toolset: Handling the diversity of english historical texts. In A. Ammermann, A. Brock, J. Pflaeging, & P. Schildhauer (Eds.), *Facets of linguistics* (pp. 161–172). Proceedings of the 14th Norddeutsches Linguistisches Kolloquium 2013. Frankfurt: Peter Lang.

Poole, D. L. & Mackworth, A. K. (2010). *Artificial intelligence: Foundations of computational agents*. Cambridge: Cambridge University Press.

Postmes, T. & Spears, R. (1998). Deindividatuation and antinormative behavior: A meta-analysis. *Psychological Bulletin*, *123*, 238–259.

Russell, S. (1997). Rationality and intelligence. *Artificial Intelligence*, *94*, 57–77.

Russell, S. & Norvig, P. (2010). *Artificial intelligence: A modern approach*. Englewood Cliffs: Prentice-Hall.

Sandholm, T. (2007). Expressive commerce and its application to sourcing: How we conducted $35 billion of generalized combinatorial auctions. *AI Magazine*, *28*, 45–58.

Schaufeli, W. B. & Bakker, A. B. (2010). The conceptualization and measurement of work engagement: A review. In A. B. Bakker & M. P. Leiter (Eds.), *Work engagement: A handbook of essential theory and research* (pp. 10–24). New York: Psychology Press.

Sonnentag, S. & Frese, M. (2002). Performance concepts and performance theory. In S. Sonnentag (Ed.), *Psychological management of individual performance* (pp. 3–25). New York: John Wiley & Sons.

Stroebe, W., Diehl, M., & Abakoumkin, G. (1996). Social compensation and the köhler effect: Toward a theoretical explanation of motivation gains in group productivity. In E. H. Witte & J. H. Davis (Eds.), *Understanding group behavior: Small group processes and interpersonal relations* (Vol. 2, pp. 37–65). Hillsdale: Erlbaum.

Stubbs, M. (2001). Computer-assisted text and corpus analysis: Lexical cohesion and communicative competence. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *Handbook of discourse analysis* (pp. 304–320). Blackwell Handbooks in Linguistics. Oxford: Blackwell.

Stürmer, S. & Snyder, M. (Eds.). (2010). *The psychology of prosocial behavior: Group prozesses, intergroup relations, and helping*. Chichester: Wiley-Blackwell.

Tajfel, H. & Turner, J. C. (2004). The social identity theory of intergroup behavior. In J. T. Jost & J. Sidanius (Eds.), *Key readings in social psychology: Political psychology* (pp. 276–293). New York: Psychology Press.

Van den Broeck, A., De Cuyper, N., De Witte, H., & Vansteenkiste, M. (2010). Not all job demands are equal: Differentiating job hindrances and job challenges in the job demandsresources model. *European Journal of Work and Organizational Psychology*, *19*, 735–759.

Williams, K. & Karau, S. (1991). Social loafing and social compensations: The effects of expectations of co-worker performance. *Journal of Personality and Social Psychology*, *61*, 570–581.

Witte, E. H. (1989). Köhler rediscovered: The anti-ringelmann effect. *European Journal of Social Psychology*, *19*, 147–154.

Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, *40*, 1–95.