Competence and Motivation in Volunteer Translators: The example of the Translation Commons

Tabea De Wille, Magister Artium
Student number: 12093645

Dissertation submitted for the degree of Master of Science
Department of Computer Science & Information Systems
University of Limerick

Supervisor:
Reinhard Schäler

September 2014
Thanks

While I am grateful to many people for their support, not least my supervisor Mr Reinhard Schäler and my parents Eberhard and Martina De Wille, I would especially like to thank Paulina Abzieher and Alison Nolan from The Rosetta Foundation for their invaluable help in gathering the data used for analysis in this paper as well as providing pointers that aided the understanding of the data and workflows on Trommons.
1: Introduction ................................................................................................................................. 4
1.1. Background .............................................................................................................................. 4
1.2. Research question .................................................................................................................... 5
1.3. Structure of the paper ............................................................................................................. 6
2: Literature and theoretical foundations ...................................................................................... 6
2.1. Review of the literature .......................................................................................................... 6
2.2. Motivation theory .................................................................................................................... 9
2.3. Why people volunteer .......................................................................................................... 14
3: Trommons - UI and features ...................................................................................................... 16
3.1. Introduction to Trommons .................................................................................................... 16
3.2. Trommons feature and workflow description ....................................................................... 17
   3.2.1. Profile and profile creation ............................................................................................ 17
   3.2.2. Task creation, search and accepting .............................................................................. 21
   3.2.3. Working on tasks, getting help and task completion ..................................................... 25
   3.2.4. Review and feedback, post-delivery ............................................................................. 29
3.3. Trommons features and workflow comments ....................................................................... 30
   3.3.1. Profile and profile creation ............................................................................................ 30
   3.3.2. Task creation, search and accepting .............................................................................. 32
   3.3.3. Working on tasks, getting help and task completion ..................................................... 34
   3.3.4. Review and feedback, post-delivery ............................................................................. 37
4: Output review and feedback .................................................................................................... 38
4.1. Composition of volunteer data ............................................................................................. 38
4.2. Coding and analysis of the comments ................................................................................... 43
   4.2.1. Coding and analysis method ......................................................................................... 43
   4.2.2. Analysis results ............................................................................................................. 51
5: Discussion and recommendations ............................................................................................ 75
Reference list .................................................................................................................................. 78
1: Introduction

1.1. Background

O’Brien & Schäler (2010) estimate that volunteer work ranks ahead of the top translation companies. "Thousands of translators have long been involved in the localization of open source success stories such as Mozilla and Open Office, providing access to basic IT tools and technologies to those who do not represent a market for large multinational companies." (O’Brien & Schäler 2010, pp. 1-2). While they also state that the volumes are not the most important aspect of the volunteer work, but instead the impact on the languages covered as well as the lives of the people they translate for (O’Brien & Schäler 2010, p. 2), it nonetheless points towards a growing importance of volunteer work in translation and localisation.

This is also supported by data on the volunteer market overall. In the US alone, "(i)n 2012, one in four adults (26.5 percent) volunteered through an organization, (...) (a)ltogether, 64.5 million Americans volunteered nearly 7.9 billion hours last year. The estimated value of this volunteer service is nearly $175 billion, based on the Independent Sector’s estimate of the average value of a volunteer hour. (...) Volunteering has trended upward among Generation Xers (born between 1965 and 1981) over the past eleven years, increasing nearly 5.5 percentage points during that period." (Corporation for National & Community Service n.d.)

One of the central questions in volunteer work of all areas is however, why "(...) people decide, in the first place, to engage in helpful activities as volunteers? And, having decided to volunteer, why do people continue to serve, sometimes for months and even years?" (Clary & Snyder 1999, p. 156) By answering those questions, it might then be possible to not only engage more people and motivate them to volunteer, but also retain them so that they continue to provide voluntary work.

In a survey of the volunteers who had signed up for projects for The Rosetta Foundation (TRF), O’Brien & Schäler (2010) conducted research into the question what motivating factors for volunteer translators, specifically at TRF, were. For this purpose, a survey was undertaken and the results presented. They also investigated what might better motivate volunteers to contribute (O’Brien & Schäler 2010, p. 1). The Rosetta Foundation is a non-profit spin-out of the Centre For Global Intelligent Content (CNGL), and "has mobilised 5,000 language volunteers worldwide on its innovative Translation Commons platform. The platform is growing by 64% annually, delivering free translation services in 88 languages to 120 non-profit organisations in 27 countries." (Centre for Global Intelligent Content 2014).

Of the motivating factors, the highly motivating ones were support of The Rosetta Foundation’s cause, followed by the wish to gain professional translation experience. The lowest were improving language skills and wishing to support lesser-used languages. "This suggests that, at least in the case of TRF, volunteer translators are motivated by both personal goals and social causes." (O’Brien & Schäler 2010, p. 6)

Of the factors that motivate future volunteering, highest ranking was feedback from both TRF clients and professional translators. This means there is a firm link between what initially motivates volunteers (wanting to gain professional translation experience and getting feedback from professional translators) and the ongoing motivational factor of getting feedback from professional translators (O’Brien & Schäler 2010, p. 7).
"It would also seem important that a mechanism for feedback on translation quality should be created. For example, such a mechanism exists for volunteer translators of Ted.com content: the translator is put in touch with a QA linguist to discuss the translation and suggested modifications, where necessary." (O’Brien & Schäler 2010, p. 9)

Since this survey of TRF volunteers, the Translation Commons platform (trommons.org) has been launched, which also includes such a mechanism for feedback on translation quality. This has motivated the research questions for this paper.

1.2. Research question

Bearing in mind the background outlined above, the main research question is:

To what extent does Trommons support volunteer motivation by satisfying their psychological need of feeling competent?

In order to answer this question, multiple angles need to be considered. One, is differentiating between the expressed goal of volunteers to gain experience and receive feedback. While only a portion of the volunteers have stated in the previously mentioned survey that this is important to them, this paper will cite sources showing that the need to feel competent is innate to every human being and that this feeling of competence can be supported or hindered with feedback. For that reason, all volunteers will be treated with the assumption that they would like to have that need satisfied, even if they are not actively trying to improve their skills or gain experience.

Two, is the question to what degree the platform itself is suited to increase rather diminish the feeling of competence while volunteers create their profile, look for tasks and accept them, work on tasks, deliver them, interact with other users and track their progress in gaining competence. This is a very multi-faceted topic, and can among other aspects include elements of user interface (UI) design (Does the user feel competent when using the platform itself or are there obstacles, confusing elements? Is it clear what is expected of the user in regards to tasks as well as interaction with the UI and features?), the actual work on the task itself (In the broadest terms, is it clear to the user, both volunteer and representative of the organisation, what actions will lead to the best outcome possible in regards to workflow, need satisfaction and quality?) as well as how much users promote or limit competence motivation of other users through the feedback system.

In order to limit the scope of this paper, only a brief analysis of the user interface and the support for working on tasks has been conducted and its results presented in this paper, albeit without consulting relevant literature, which means that this analysis part is likely to be subjective and will require further research.

The main part of this paper is the analysis of the feedback system, with its numeric rating and the comment fields. Here the questions of how users are using the numeric star rating system, how the
numeric star rating relates to the comments that were left and what the nature of the comments is were investigated against the background of Self-determination theory and volunteer motivation.

1.3. Structure of the paper

This paper is structured to first include a review of the relevant literature in the area of motivation in community translation, followed by an introduction of the theoretical background applied to the research. The next section is an overview of the main features on Trommons and an introduction to the workflow with a focus on the volunteer perspective, followed by impressions and suggestions on the UI and features, with the cited literature in mind.

The third part introduces the method of analysing the data provided by TRF which included numeric feedback as well as written comments provided by volunteers and organisations on Trommons. It then shows the results of the analysis, and concludes with a discussion of the outcomes and offers some suggestions.

2: Literature and theoretical foundations

2.1. Review of the literature

Only limited research in the area of volunteer translations in relation to motivations has been conducted to date. This was also observed by Dombek (2012) who notes that "(...) the literature search on motivation uncovered almost no research that is applicable to the current discussion on professional and volunteer translator motivation in the field of Translation Studies" (Dombek 2012, p. 13). "Nevertheless, through analyzing the behavioral manifestations of motivation (such as effort, persistence, personal investment, physiology and self-report) it is possible to deduce and ascribe to a particular person his or her internal motives - needs, cognitions and emotions that energize and direct behavior." (Dombek 2012, p. 14)

Dombek (2012) investigates why volunteers on Facebook contribute their free translations (Dombek 2012, p. 1), while Facebook doesn't offer financial remuneration but rewards its users with "awards" and listing people on its leader board (Dombek 2012, p. 12).

While the users of both, Facebook and Trommons volunteer their translation efforts, there are also some significant differences as the overview of the Trommons features will show. Facebook segments its content and displays it to the user in random order. It also offers an "inline" translation mode where users can translate segments while using Facebook for normal activities (Dombek 2012, p. 9) and volunteers can vote on the quality provided by others. They can use up or down voting and there are values added based on the reputation of the person voting. On the basis of these votes, the highest score is selected and all translations are reviewed by professional translators (Dombek 2012, p. 10). This is especially interesting since the way Facebook segments content and displays content randomly in its
application with a segment usually not being related to the previous one (Dombek 2012, p. 8), it could be expected that volunteer motivation is low due to issues with autonomy according to Self-determination theory (Ryan & Deci 2000). In addition to that, the feedback system of voting translations up and down doesn't immediately appear suited to help volunteers improve their skills.

However, Facebook also provides other functions to support volunteers like a discussion board, glossaries and stylistic guidelines (Dombek 2012, p. 11)

The Facebook platform, like other crowdsourcing platforms, not only facilitates the translation process, but also "(...) must function in a way that lets their users achieve all the outcomes they expect to achieve by joining the crowd and getting involved in a particular crowdsourcing initiative. Consequently, when these expectations are met, they continue with their efforts and keep contributing." (Dombek 2012, p. 18)

Dombek (2014) finds that there is a variety of reasons that motivate users to contribute to Facebook crowdsourcing and employs four theoretical frameworks: self-determination theory (SDT), functional approach to volunteer motivation, motivation to collaborate online and gamification (Dombek 2014, p. 260). Motivating factors can be correspondingly attributable when users are provided with an opportunity to satisfy the needs of competence, autonomy and relatedness, when their expectations on personal and social benefits of their voluntary activity are met, when they experience reciprocity, self-efficacy, group commitment and reputation gain or when the activity is perceived as fun and enjoyable, undertaken for pleasure (Dombek 2014, p. 260).

Dombek finds that the Facebook translation platform satisfies a lot of these needs, in particular by challenging the users' skills, enabling further development of their linguistic knowledge and translation abilities, generating social capital and benefiting the public. Users can monitor their own progress and signal their achievements to others (Dombek 2014, p. 261)

However, Dombek also identifies that the translation platform sometimes became a source of frustration and discontent. The source for this could be found in the translation platform preventing users from providing appropriate, correct or natural translations. It thereby was found to weaken the need satisfaction, diminish the opportunity to meet expectations, hindering reciprocity, self-efficacy, group commitment and reputation gain as well as limiting the perception of fun for the users (Dombek 2014, p. 262). She therefore concludes that the choice of technology, its design and functionality need to be carefully considered and will not only affect the quality of translation by the crowd but also the retention rate of the contributing individuals (Dombek 2014, p. 271). This conclusion can also applied to the Trommons platform since its design features limit human intervention in the workflow, which means that it also limits the opportunities for organisations or TRF staff to increase motivation and smooth out factors that decrease motivation, especially if the user does not initiate a dialog about issues they might encounter.

Kelly et al. (2011) compare the traditional translation-editing-proofreading (TEP) model with aspects of community translation or crowdsourcing. They discuss that the traditional TEP model inhibits collaboration and its drawbacks on quality and efficiency, especially related to information sharing and timing (Kelly et al. 2011, pp. 75-6). "Perhaps the greatest drawback of the TEP model is that it only
enables individuals to detect errors at certain checkpoints or at the end of a project." (Kelly et al. 2011, p. 76)

In contrast, they describe a crowdsourcing model as an example of a collaborative method where project teams aren't primarily organized by language as with the TEP model, but instead would include many teams working on different languages, allowing participants to interact across language groups (Kelly et al. 2011, p. 77). Additionally, it allows for real-time interaction with subject matter experts who would vet translations, ask questions and suggest improvements (Kelly et al. 2011, p. 78). "Through this collaborative process, the editing stage would be eliminated, since quality improvement would take place at essentially the same time as translation." (Kelly et al. 2011, p. 78)

In this model, the translator is an active member of a community, entering questions in an online group rather than looking up phrases or words in a dictionary, asking follow-up questions and submitting queries for specific contexts or exceptions. The translator would also be able to review other people's translation memory files (Kelly et al. 2011, p. 79). In addition, in such a model, "(...) human subject matter experts coexist and contribute to high-quality translations in the first pass. In fact, in a collaborative model, subject matter experts do not just do a final review at the end, but rather, are relied upon throughout the process to answer questions in a timely manner so that translators are not left waiting. Translators, thus, no longer rely on the editor's eyes to catch their mistakes at the end, and must be trained to ask questions to help them translate correctly from the start." (Kelly et al. 2011, p. 80)

Thereby, the focus is not on fixing errors in translations but on creating correct translations from the start through feedback and guidance from subject matter experts throughout the process (Kelly et al. 2011, p. 80). In the collaborative model, translators would also have the ability to collaborate and learn from each other cross language pairs (Kelly et al. 2011, p. 81). "Instead of reporting issues to the project manager as happens in the TEP model, translators in the collaborative model tag and insert comments in real time in the original document, thus helping translators in other languages who might otherwise face the same difficulties." (Kelly et al. 2011, p. 83)

The case for a strong collaboration element in volunteer translation is also supported by Dombek who observes that "(...) many aspects of the Facebook translation crowdsourcing positively affecting the motivation of the user-translators are associated with the fact that translation crowdsourcing is inherently an activity performed collaboratively online." (Dombek 2014, p. 263)

Due to the limited previous research in the area, a broader approach to the topic had to be taken. This includes Self-determination Theory, as well as competence and achievement motivation with some elements of the impact of feedback on perceived competence.

The second part of the theoretical background includes literature dealing with volunteerism and altruism on a more general level and not specific to translation and localisation.
2.2. Motivation theory

Self-Determination Theory (SDT) as described by Ryan & Deci (2000) includes factors that can lead to humans being proactive and engaged or passive and alienated, which means they enhance or undermine intrinsic motivation. They postulate that there are three psychological needs that are innate for all people: competence, autonomy and relatedness (Ryan & Deci 2000, p. 68). These psychological needs "appear to be essential for facilitating optimal functioning of the natural propensities for growth and integration, as well as for constructive social development and personal well-being." (Ryan & Deci 2000, p. 68). This observation supports the approach of considering the need to feel competent satisfied universal to all volunteers on Trommons, in addition to the group who declared the wish to gain experience and receive feedback as a reason to volunteer.

There are also factors that hinder or undermine self-motivation, social functioning and personal well-being. These can be described as thwarting the three basic psychological needs. SDT is therefore not only concerned with the positive development tendencies but also environments that hinder these tendencies (Ryan & Deci 2000, p. 69). By considering whether a factor supports or thwarts the satisfaction of the three psychological needs mentioned, people can predict effects of that factor on outcomes like motivation, behaviour, affect and well-being (Deci & Ryan 2008, p. 183).

Humans tend to be intrinsically motivated, but this tendency requires supportive conditions since it can be disrupted by non-supportive conditions (Ryan & Deci 2000, p. 70). "Thus, our theory of intrinsic motivation does not concern what causes intrinsic motivation (which we view as an evolved propensity; [...]); rather, it examines the conditions that elicit and sustain, versus subdue and diminish, this innate propensity (Ryan & Deci 2000, p. 70).

Extrinsic rewards can undermine intrinsic motivation since it facilitates a more external perceived locus of causality. Therefore, tangible rewards related to task performance reliably undermine intrinsic motivation. The same is the case for threats, deadlines, directives, pressured evaluations, and imposed goals (Ryan & Deci 2000, p. 70). "In contrast, choice, acknowledgment of feelings, and opportunities for self-direction were found to enhance intrinsic motivation because they allow people a greater feeling of autonomy." (Ryan & Deci 2000, p. 70)

While intrinsic motivation is an important type of motivation, it is not the only type/the only type of self-determined motivation. Strictly speaking, after early childhood a lot of what people do isn't intrinsically motivated due to social pressures to do activities that are not interesting and to assume responsibilities (Ryan & Deci 2000, p. 71).

SDT distinguishes between autonomous motivation and controlled motivation:

- Autonomous motivation: intrinsic motivation as well as the types of extrinsic motivation that are identified or integrated (people have identified with an activity's value and ideally have integrated it into their sense of self) (Deci & Ryan 2008, p. 182)

- Controlled motivation: consists of external regulation (behaviour is a function of external contingencies of reward or punishment) as well as introjected regulation (regulation of action has been partially internalized; energized by approval motive, avoidance of shame, contingent self-esteem and ego-involvements) (Deci & Ryan 2008, p. 182)
SDT has addressed the issues of how behaviours that are nonintrinsically motivated can become truly self-determined and how the social environment influences those processes. "The term extrinsic motivation refers to the performance of an activity in order to attain some separable outcome and, thus, contrast with intrinsic motivation, which refers to doing an activity for the inherent satisfaction of the activity itself." (Ryan & Deci 2000, p. 71)

However, "(...) SDT proposes that extrinsic motivation can vary greatly in its relative autonomy." (Ryan & Deci 2000, p. 71). Extrinsically motivated behaviour can entail personal endorsement and a feeling of choice, or compliance with an external regulation (Ryan & Deci 2000, p. 71)

Ryan & Deci describe this as a Self-Determination Continuum with types of motivation with their regulatory styles, loci of causality, and corresponding processes (Ryan & Deci 2000, p. 72).

The continuum reaches from amotivation on the left (the state of lacking the intention to act, where people act without intent and just go through the motions; they don’t value an activity and don’t feel competent to do it or don’t expect it to yield a desired outcome) (Ryan & Deci 2000, p. 72) to the classic state of intrinsic motivation, which describes doing an activity for its inherent satisfaction. "It is highly autonomous and represents the prototypic instance of self-determination." (Ryan & Deci 2000, p. 72)

The continuum includes extrinsically motivated behaviours which reach from least autonomous (externally regulated; performed to satisfy an external demand) via introjected regulation (taking in a regulation but not fully accepting it as one’s own, where "behaviors are performed to avoid guilt or anxiety or to attain ego enhancement such as pride" (Ryan & Deci 2000, p. 72)). Here people demonstrate the ability to maintain feelings of worth. They are internally driven but still have an external perceived locus of causality (Ryan & Deci 2000, p. 72).

The next group towards a more autonomous form of extrinsic motivation is "regulation through identification". It reflects conscious valuing of a behavioural goal or regulation. The action is accepted as personally important (Ryan & Deci 2000, p. 72).

The most autonomous form of extrinsic motivation is "integrated regulation". It occurs when regulations are fully assimilated to the self, they have been evaluated and brought into congruence with one’s other values and needs. "Actions characterized by integrated motivation share many qualities with intrinsic motivation, although they are still considered extrinsic because they are done to attain separable outcomes rather than for their inherent enjoyment." (Ryan & Deci 2000, p. 73)

When applied to the volunteers on Trommons, this suggests that while they might not always be intrinsically motivated to do certain tasks or process steps, since we can’t assume that the act of translating a document or using a spell checking program on it is inherently pleasurable for all volunteers, there are nonetheless different levels of autonomy that can be achieved within the continuum of extrinsic motivation. A poor outcome for example would be if volunteers were to comply with certain quality standards or complete tasks for fear of punishment, like being publicly shamed if their delivery was poor or being pressured into delivering a task otherwise. Ideally, volunteers would reach at least the highest level of autonomy in extrinsic motivation where they have integrated a certain level of quality checks, a sequence of actions in a workflow or the overall notion of why volunteering on Trommons is in line with their values and desires.
Initially, people perform extrinsically motivated actions because the behaviour is prompted, modelled or valued by someone they feel or want to feel attached or related to. "This suggests that relatedness, the need to feel belongingness and connectedness with others, is centrally important to internalization." (Ryan & Deci 2000, p. 73) This underlines the importance of the community element in volunteer translation, since without human interaction or an awareness of the behaviour of others within the community, people are unable to model their behaviour in accordance with the values of the community. In fact, a lack of community interaction raises the question whether there is a widely accepted set of values and behaviours if there is no synchronization process of ideas and opinions.

"The relative internalization of extrinsically motivated activities is also a function of perceived competence. People are more likely to adopt activities that relevant social groups value when they feel efficacious with respect to those activities." (Ryan & Deci 2000, p. 73) Therefore, supports for competence should facilitate internalization for the volunteers on Trommons. This suggests that by supporting the feeling of competence, among the other factors described in SDT (autonomy and relatedness), volunteers on Trommons could move towards integrated regulation, even if they are lacking intrinsic motivation for a task. Since motivation can also be thwarted on the other hand, factors that decrease integration by decreasing the feeling of competence should be avoided.

However, in order to be able to evaluate factors that may support the feeling of competence, a working definition of competence is necessary. For that, Elliot & Dweck (2005) refer to the dictionary definition. "Based on Webster’s Revised Unabridged Dictionary and the Oxford English Dictionary, 'competence' may be defined as a condition or quality of effectiveness, ability, sufficiency, or success." (Elliot & Dweck 2005, p. 5)

Elliot & Dweck also ask how competence is evaluated, to what levels of action and domains it applies and how individuals are motivated with regard to competence (Elliot & Dweck 2005, p. 5). According to them, competence evaluation is possible by the absolute standard inherent in a task, interpersonal standards implicating change over time and interpersonal standards implicating normative comparison (Elliot & Dweck 2005, pp. 5-6). Depending on how the competence is evaluated, it influences the psychological meaning that competence has (Elliot & Dweck 2005, p. 6). When analysing competence in the context of motivation, the ways in which behaviour is energized (instigated, activated) and the way it is directed (focused, aimed) needs to be accounted for. Since the need for competence is considered a fundamental motivation, the "... need for competence instigates and activates behavior that is oriented towards competence." (Elliot & Dweck 2005, p. 6). It is also possible for people to develop behaviour that is aimed at avoiding the feeling of incompetence, which could serve as a self-protective function but often doesn't lead to positive competence outcomes and experience required for continued growth and development (Elliot & Dweck 2005, p. 6). "Using a dictionary, 'competence' may be defined in purely appetitive fashion with regard to effectiveness, ability, sufficiency and success but from a motivational standpoint, the study of competence-relevant motivation will necessarily entail consideration of ineffectiveness, inability, insufficiency and failure as well." (Elliot & Dweck 2005, p. 6)

When considering volunteer translations on Trommons, this implies that while there are certain absolute standards that a volunteer might assume as inherent in a task and some normative comparisons possible, for example when a volunteer has previous experience with working on projects outside of
Trommons and has internalized the standards that were provided for such projects, due to a lack of explicit standards and expectations provided of Trommons, the interpersonal standards implicating change over time will probably play a greater role. As Elliot & Dweck (2005) show, it is also important to not only consider the aspects that can motivate volunteers, like things that are positive in the UI design and set of features on Trommons or feedback that increases their feeling of competence, but also those elements in the UI, features and feedback that are suited to diminish their perceived competence.

Sternberg (2005) postulates that there is a continuum on which people develop abilities, competencies and expertise (Sternberg 2005, p. 16). He defines developing competence as "(...) the ongoing process of the acquisition and consolidation of a set of skills needed for performance in one or more life domains at the journeyman-level or above" (Sternberg 2005, p. 15), developing expertise as "(...) the ongoing process of the acquisition and consolidation of a set of skills needed for a high level of mastery in one or more domains of life performance." (Sternberg 2005, p. 15) and experts as people who have developed their competencies to a high level (Sternberg 2005, p. 15). On the other hand, competent individuals are people who have developed their abilities to a high level (Sternberg 2005, pp. 15-6).

This is likely to be reflected in the volunteer demographics with some volunteers being in the process of developing their abilities, others being competent and again others experts. This is also confirmed in O'Brien & Schäler (2010), whose survey data showed that 222 volunteers signed up as "professional translators", while 35 classified themselves in the "amateurs" category. (O'Brien & Schäler 2010, p. 4) "Volunteers were asked to indicate their level of relevant work experience and they could chose (sic) from categories 'beginner', 'some', 'a bit out of practice', 'a lot'. 42% indicated that they had 'a lot' of experience, while 18% indicated they were 'beginners' (even though they had signed up as professional translators). Only 2% stated they were a bit out of practice, which leaves 38% as having 'some' experience. (...) Those who signed up as amateurs were also asked to state if they had experience, some experience or no experience with translation. Only 23% said they had no experience, while 46% said they had some." (O'Brien & Schäler 2010, p. 5)

In addition to the different experience levels, the differing ways people react to challenges and risks need to be considered. According to Sternberg (Sternberg 2005, p. 19) one of the different kinds of motivation is achievement motivation. "People who are high in achievement motivation seek moderate challenges and risks. They are attracted to tasks that are neither very easy nor very hard. They are strivers - constantly trying to better themselves and their accomplishments. A second kind of motivation is competence (self-efficacy) motivation, which refers to a persons' beliefs in their own ability to solve the problem at hand (...)"(Sternberg 2005, p. 19)Some people believe that to be smart is to show that one is smart, making no mistakes or showing intellectual weakness otherwise (Sternberg 2005, p. 19). "Incremental theorists, in contrast, believe that to be smart is to learn and to increase one's intellectual skills. These individuals are not afraid to make mistakes and even believe that making mistakes can be useful, because it is a way to learn." (Sternberg 2005, p. 19)

Under conditions of challenge, incremental theorists perform better since they are willing to take on difficult challenges, seek mastery of new and difficult material (Sternberg 2005, p. 19).

Elliot & Dweck (2005)however argue that the concept of achievement is not clearly defined and there is no broadly articulated, consensually shared understanding of the concept of 'achievement' (Elliot & Dweck 2005, p. 4), while "(...) most research in the achievement motivation literature has emerged from
Western, individualistic societies that tend to conceive of achievement in terms of individual, self-defining accomplishment in the prototypical domains of school, sports, and work." (Elliot & Dweck 2005, p. 4)

While this paper won’t go further into achievement motivation and to what extent this distinction is relevant, the concept that there may be people who react differently to challenges and the level of difficulty they are presented with when facing a task is nonetheless an important hypothesis to consider. In practical terms it may mean that some volunteers can deal with a higher level of frustration in regards to dealing with difficult tasks, or when being presented with UI elements or features they don’t understand than others. In order to minimize the risk of deterring those volunteers who struggle to deal with greater challenges, Trommons should ideally accommodate this group, even if others might not be deterred by the same obstacles.

Another aspect especially in a diverse setting as the user base on Trommons are different perceptions of expertise depending on culture. According to Sternberg (2005), cultures value expertise differently. He names the example of Latin parents who value social expertise higher than Asian or Anglo parents (Sternberg 2005, p. 28). However, while this topic is very relevant and worth investigating, it is a variable that is too big to consider for the scope of this paper. Therefore, unfortunately the analysis will have to be western-centric, likely with a certain degree of German influence since that is the cultural background of the author.

Finally, the question on how the feeling of competence and feedback interact and whether motivation can be influenced by performance feedback needs to be considered. In a study by Vallerand & Reid (1984) on the effects of verbal feedback on intrinsic motivation and whether those are mediated by perceived competence (Vallerand & Reid 1984, p. 94), they cite other studies, in which "(...) it is generally found that positive performance information increases intrinsic motivation (...) while negative performance information decreases intrinsic motivation ..." (Vallerand & Reid 1984, p. 95) They then go on to show that indeed "positive and negative feedback increases and decreases intrinsic motivation, respectively, and that these changes in intrinsic motivation are actually caused by changes in perceived competence." (Vallerand & Reid 1984, p. 99)

When applied to Trommons, this suggests that positive comments a volunteer receives as feedback on their task should increase their perceived competence and thereby motivation, while negative comments might decrease them. However, it cannot be automatically assumed that all negative comments need to be avoided, since pointing out faulty elements in a translation can help with avoiding such issues in the future. It is therefore important to determine how the criticism is worded and whether it is voiced in general terms or is specific enough to make volunteers aware of which areas they can improve on.
2.3. Why people volunteer

The question of why people volunteer their time and efforts with no or little compensation is one that has been asked also in other areas where volunteerism is prevalent, as well as on a more general level.

Clary & Snyder (1999) ask why people engage in volunteering and identify six personal and social functions that are served by volunteering. They explore the role motivation plays in the decision of becoming a volunteer, as well as the decision to continue (Clary & Snyder 1999, p. 156).

According to them, volunteering has the following characteristics: The volunteer must find opportunities to volunteer, decide, after deliberation, to help and then decide whether or not to continue, influenced by whether the activity fits with the needs and goals of the volunteer. "The effortful, sustained and nonremunerative nature of volunteering raises two fundamental questions: Why do people decide, in the first place, to engage in helpful activities as volunteers? And, having decided to volunteer, why do people continue to serve, sometimes for months and even years?" (Clary & Snyder 1999, p. 156).

Clary & Snyder (1999) cite the Volunteer Functions Inventory (VFI), which includes the functions values, understanding, enhancement, career, social and protective (Clary & Snyder 1999, p. 157). For competence as a factor influencing motivation, especially understanding and career seem relevant.

The conceptual definition of understanding is: "The volunteer is seeking to learn more about the world or exercise skills that are often unused." And the Sample VFI item: "Volunteering lets me learn through direct, hands-on experience." For career, the conceptual definition is: "The volunteer has the goal of gaining career-related experience through volunteering" and the sample VFI item: "Volunteering can help me to get my foot in the door at a place where I would like to work." (Clary & Snyder 1999, p. 157)

They find that the most common functions are Values, Understanding and Enhancement, while Career, Social and Protective are less important. However, the ordering and absolute importance vary across groups (Clary & Snyder 1999, p. 157). Additionally, the same volunteer may be pursuing more than one goal. This is reported by roughly two thirds of the respondents (Clary & Snyder 1999, p. 157).

Assuming the goal is to motivate people to become volunteers more often and remain dedicated to the project they have selected, persuasive messages that are tailored to the specific motivations important to the individual, can motivate people to initiate volunteer service (Clary & Snyder 1999, p. 158). Therefore, "(...) attempts to recruit volunteers will succeed to the extent that they address the specific motivational functions underlying behavior and attitudes." (Clary & Snyder 1999, p. 158).

If the motivational concerns are served by volunteering, the findings suggest that they would derive greater satisfaction than for those where this isn't the case (Clary & Snyder 1999, p. 158). So "... it follows that their actual intentions to continue serving as volunteers will also be linked to the matching between experiences and motivations." (Clary & Snyder 1999, p. 158)

Schultinik et al. (n.d.) describe aspects that motivate people to volunteer and how this information applies to the work of a volunteer manager (Schultinik et al. n.d., p. 1).
They cite The Independent Sector (1997) who identifies eight reasons why people volunteer:

1. To make a difference
2. To use a talent or skill.
3. To gain professional experience or make contacts.
4. To express religious faith.
5. To meet people.
6. To achieve personal growth and enhanced self-esteem.
7. To seek a more balanced life.
8. To give something back.

(Schultinik et al. n.d., p. 1)

They also cite Maslow's hierarchy of needs with physiological needs (food, shelter, clothing, water etc.), the need for safety, the need for love and belonging, the need for self and group esteem and the need for self-actualization (Schultinik et al. n.d., pp. 1-2) as well as Tjosvold's premise of everyone being driven by five forces: Biological, the need to love/be loved, the need for power/self esteem, the need for freedom/self esteem, the need for creativity/fun (Schultinik et al. n.d., p. 2).

Schultinik et al. (n.d.) then group Maslow's Hierarchy and Tjosvold's premise as well as the categories provided by the Independent Sector into the areas Achievement, Affiliation and Power. Achievement satisfies the need to be challenged as well as to set personal goals and challenges which build self-esteem.

Affiliation provides a sense of roots, belonging somewhere (place or group) as well as following the goals of the organization, being social and supportive.

Power satisfies the need for personal power, mastery over others and wanting to make a difference (Schultinik et al. n.d., p. 2). These factors are important since in order to be effective at managing and supporting volunteers, their motivations need to be understood (Schultinik et al. n.d., p. 3).

The notion that by supporting volunteers and satisfying their psychological needs, they tend to volunteer more often and longer, is also supported by Bang & Ross (2009) who have conducted a study to examine the impact of motivations on the satisfaction of individual volunteers, specifically in the field of sporting events (Bang & Ross 2009, p. 61). They show some examples of altruistic reasons why people volunteer, but also that "Shibli, et al. (1999) suggested that although a number of participants volunteered in a strong altruistic manner, the primary reason for volunteering was not altruistic, but rather to meet the needs of the volunteer." (Bang & Ross 2009, p. 62)

Bang & Ross (2009) find in their study that satisfying the psychological needs of an individual promotes further motivation (Bang & Ross 2009, p. 70). In addition to that, "... volunteers tended to be more satisfied when they felt that they helped make the event a success (...) When volunteers feel that they
are truly needed and responsible for the event, their job performance may be improved." (Bang & Ross 2009, p. 71)

This raises the question of what the potential consequences are if the needs of volunteers on Trommons are not satisfied, be they openly stated needs like the wish to improve their skills and gain experience, or the more covert need of feeling competent. Unless there are motivational factors that outweigh this, it seems likely that such volunteers would not continue to volunteer on Trommons, which could in the worst case mean that not only one volunteer is lost from the potential group of people who accept tasks, but in addition those that might have been recruited through word of mouth by them. It might also lead to volunteers abandoning a project rather delivering what they have committed to by claiming a task, or alternatively delivering poor quality. However, the reasons why people abandon tasks or to what extent they deliver poor quality and why is outside the scope of this paper, but could provide directions for further research.

3: Trommons - UI and features

3.1. Introduction to Trommons

Trommons (Translation Commons) is a web-based platform that connects large pools of volunteer translators with communities that require their services free of charge (The Rosetta Foundation 2014).

Trommons was launched in May 2013 (Centre for Global Intelligent Content 2013). However, The Rosetta Foundation, which has been given the exclusive license of Trommons (The Rosetta Foundation 2014) and manages it, had already connected volunteer translators with non-profit organizations requiring localisation services before that date. The work prior to the launch of Trommons is not subject of this paper.

According to the statistics on Trommons.org, as of August 11th 2014 Trommons has 5970 registered users, who have worked on 597 projects with 2151 claimed tasks for 144 organisations.

Trommons uses SOLAS (Service Oriented Localisation Architecture Solution), which was developed at the Localisation Research Centre at the University of Limerick. It is a translation management platform, based on the principles of flexible, open standards-, components- and web services-based development with the aim of supporting collaborative and distributed localisation workflows (Centre for Global Intelligent Content 2013).
3.2. Trommons feature and workflow description

3.2.1. Profile and profile creation

Trommons connects volunteers with organisations requiring translations by allowing both to sign up for a free account.

Volunteers have the option of providing information like their language combinations (native language and secondary languages), biography, job title and personal details. They also have the opportunity to subscribe to tags which the organisations have assigned to tasks. In addition to that, they can indicate whether they are interested in translation, proofreading or interpreting tasks.

On their profile, volunteers can sign up for task stream notifications. Those are e-mails that are sent out from Trommons if a task matches with the volunteer’s preferences. (Image 1)

The volunteer profile also includes a section with badges which can be created and awarded by the organisations, as well as the option to request a reference e-mail. The mail appears to be automated and includes a list of tasks the volunteer has contributed to, as well as a note of appreciation from The Rosetta Foundation CEO Reinhard Schäler. (Image 2)

Volunteers can also see a list of tasks they have worked on in the past, which includes those tasks that have since been archived. (Image 3)

Organisations can include details like their biography, location, name, home page and e-mail, as well as regional focus on their biography. However, only the name is required. Once the profile has been created, organisations are also able to create badges and add members to their organisation.
Image 1: Sample task stream notification

Special Olympics_Program Quality Standards

Type: Proofreading
From: English
To: German
Tags: korean dutch German serbian farsi
Word Count: 4000
Added: 12 May 2014 - 21:06
Due by: 30 December 2014 - 00:00
Part of: Special Olympics_Program Quality Standards for Special Olympics
Image 2: Sample reference letter

On behalf of The Rosetta Foundation and especially our partner organisations that have benefited from your work, I would like to thank you for your dedication and enthusiastic support in removing the price tag for access to information and knowledge across languages for those who need it most.

Below, you will find a record of your work with The Rosetta Foundation. We hope that you found volunteering your skills rewarding and we look forward to continue working with you in future.

Please do not hesitate to contact us with any queries or suggestions you may have.

Kind Regards,
Reinhard Schaler
CEO

Tasks You Contributed To

<table>
<thead>
<tr>
<th>Task Title</th>
<th>Organisation</th>
<th>Task Type</th>
<th>Source</th>
<th>Target</th>
<th>Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Troika song</td>
<td>International Citizen Audit Network (ICAN)</td>
<td>Proofreading</td>
<td>English (UNITED KINGDOM)</td>
<td>German (GERMANY)</td>
<td>319</td>
</tr>
<tr>
<td>Competitiveness and The Troika (DE &amp; IT)</td>
<td>International Citizen Audit Network (ICAN)</td>
<td>Translation</td>
<td>English (UNITED KINGDOM)</td>
<td>German (GERMANY)</td>
<td>2340</td>
</tr>
<tr>
<td>TTIP/TAFTA (DE &amp; IT)</td>
<td>International Citizen Audit Network (ICAN)</td>
<td>Translation</td>
<td>English (UNITED KINGDOM)</td>
<td>German (GERMANY)</td>
<td>1596</td>
</tr>
</tbody>
</table>
Image 3: Sample archived task

Archived Tasks A list of tasks you have worked on in the past.

Troika song
Task Type: Proofreading
From English - UNITED KINGDOM (en - GB)
To German - GERMANY (de - DE)
Word Count: 319
Added 6 months, 1 week ago
Archived 4 months, 1 week ago
Due by 12 February 2014 - 00:00 UTC
3.2.2. Task creation, search and accepting

When creating a project, organisations fill in the source and target language or languages, the tasks that need to be done, as well as the task deadline. They name the project and provide a project description as well as having the option of describing the impact the project has and adding tags for the project. They are then notified when a volunteer accepts, comments on or delivers the task.

Volunteers can search for tasks using the criteria source language, target language and task type (segmentation, desegmentation, translation and proofreading). Individual tasks can carry the information project name, source and target language, task creation, task deadline, word count, task comment, project description, project impact and project tags. (Image 4)

Additionally, there is a document preview available that allows volunteers to read the content prior to deciding whether they want to work on it. It is also possible to download the document.

When clicking on the project name, another view becomes available which contains some of the information already provided again (source language, word count, creation date, deadline, project description, project impact and tags). In addition, the organisation name linking to the organisation profile, as well as a reference field is available on this screen. (Image 5)

After a volunteer downloads the file, they see a confirmation screen (Image 6) which asks them to ensure they can open the file, whether they have enough time and the capability to work on it and provides a link to the original source file in case of a proofreading task. The volunteer is then asked to promise that they will complete the task.

Once the volunteer has accepted the task, they are instructed to check whether they can open the file, that they should finish the task and then upload it to Trommons again. (Image 7)
<table>
<thead>
<tr>
<th>Project</th>
<th>Source Language</th>
<th>Target Language</th>
<th>Created</th>
<th>Task Deadline</th>
<th>Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indigenous healthcare project in the rainforest of Mexico - translation of our website.</td>
<td>Spanish - MEXICO (es - MX)</td>
<td>German - GERMANY (de - DE)</td>
<td>8 August 2014 - 18:16 UTC</td>
<td>27 August 2014 - 00:00 UTC</td>
<td>1800</td>
</tr>
</tbody>
</table>

**Task Comment:**
No comment has been listed.

**Project Description:**
The Collective Tzune le Bel ("Planting the Way" in the local Mayan Tzeltal language) works to support self-determined health in the Lacandon Rainforest of Chiapas in South Mexico. We would like to translate our website at http://tunnelbel.webify.com/ into English and German. We please note that I have only uploaded the front page here, but we would like to translate the whole website (all pages).

**Project Impact:**
The translation of our page will help us get more support from friends and individuals around the world who like our work in the communities - and who don’t necessarily speak Spanish.

**Project Tags:**
- health
- non-profit
- indigenous
- mexico
- rainforest
- chiapas
- organic-agriculture
- communities
**Indigenous healthcare project in the rainforest of Mexico - translation of our website.** Overview of project details.

<table>
<thead>
<tr>
<th>Organisation</th>
<th>Source Language</th>
<th>Reference</th>
<th>Word Count</th>
<th>Created</th>
<th>Project Deadline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tunel Bej</td>
<td>Spanish - MEXICO (es-MX)</td>
<td><a href="http://tunelbej.weebly.com">http://tunelbej.weebly.com</a></td>
<td>1800</td>
<td>8 August 2014 - 18:16 UTC</td>
<td>31 August 2014 - 00:00 UTC</td>
</tr>
</tbody>
</table>

**Description**

The Collective Tunel Bej ("Planting the Way"") in the local Mayan Tzeltal language works to support self-determined health in the Lacandon Rainforest of Chiapas in South Mexico. In ViViWo we would like to translate our website at http://tunelbej.weebly.com/ into English and German. ViViWo! Please note that I have only uploaded the front page here, but we would like to translate the whole website (all pages).

**Impact**

The translation of our page will help us get more support from friends and individuals around the world who like our work in the communities - and who don't necessarily speak Spanish.

**Tags**

- health
- eco-profit
- indigenous
- mexico
- rainforest
- chiapas
- organic-agriculture
- communities
**Do you want to proofread this file?** After downloading

**Review this checklist for your downloaded file.** Will you be able to proofread this file?

1. Can you open the file on your computer?
2. Will you have enough time to proofread this file? Check how long the file is.
3. Do you think you are capable of proofreading a file in Abai (abiw)?
4. If you would like to view the original file in its source language you can download it [here](#).

[Download File]

**It is time to decide**

Do you want to proofread this file? When you are finished proofreading the file, you will need to upload it again.

- Yes, I promise I will proofread this file
- No, just bring me back to the task page

---

**Image 7: Sample instructions**

**What happens now?** We need your translation.

This is what you need to do (as soon as possible)

1. Can you open the file on your computer?
2. Translate the file to Abai Sungai (abf) using your favourite translation software.
3. Upload your finished translated file to the task page.

We have also emailed you these instructions to [tabea.dewille@gmail.com](mailto:tabea.dewille@gmail.com).

---

**When you have finished translating the file you downloaded:**

- Upload Translated Task
- No, just bring me back to the task page.
3.2.3. Working on tasks, getting help and task completion

After the volunteer has claimed a task, they are able to send a message to the organisation and provide them with feedback about the task. They are also able to find source files and the latest version of the file related to their task, as well as the information regarding the task and the organisation that had already been available prior to accepting the task.

The volunteer is also able to unclaim a task if they feel they can't complete a task after having claimed it.

While volunteers are working on tasks, they are able to refer to the forums at forum.trommons.org - the sections available are Welcome, Volunteers Q&A, Non-profits Q&A, Projects and Resources.

At time of writing, there is a total number of 8 topics with a total number of 19 posts and 18 registered users. Users are required to register for the forums separately from their Trommons account if they wish to post comments, but are able to read the content without having registered or being logged in.

Trommons.org also includes a section with frequently asked questions, where volunteers can get help with technical questions specific to Trommons and its workflow. In addition, there is one section that provides information on who to contact with queries. (Image 8)

Image 8: Volunteer FAQ list

Volunteer Questions

- How do I download the original file?
- Who do I contact with queries? Who’s the PM?
- What if I need more time?
- How do I unclaim a task?
- I am having trouble downloading the file. What do I do?
- Why does my language not appear on Trommons?
- There are no tasks for my language pair(s).
- How do I view my task reviews?
- How do I request a reference email?
- How do I create an avatar on my profile page?
- I have finished proofreading the translation. Do I upload a clean file or one with track changes?
- Where can I find out more about The Rosetta Foundation?
- How can I tell which tasks I’ve uploaded?
The question "Who do I contact with queries? Who's the PM?" provides the answer: "If you have any queries, please contact info@trommons.org. The aim of Trommons is to reduce intervention levels. Therefore, there are no Project Managers assigned to projects. Any queries can be submitted to the email previously mentioned." (The Rosetta Foundation n.d.)

The same FAQ page also includes a section for organisations (Image 9), which apart from technical questions also includes one question on feedback. The question "How do I give feedback?" is answered with "You can find a box for feedback at the bottom of the download page. There are 4 categories for feedback ranked from 1 star to 5 stars for corrections, grammar, spelling and consistency. There is also a box where any additional comments can be added." (The Rosetta Foundation n.d.)

**Image 9: Organisation FAQ list**

- How do I create an organisation?
- How do I download my translation?
- How do I know when my project is finished?
- How do I add more members to my organisation?
- How do I give feedback?
- How do I view task reviews on my projects?
- My project file is quite large. What do I do?

The third section on the FAQ page includes common terms within Solas Match. Those questions explain features of the user interface on Trommons, and clarify what a task, segmentation and project impact include. "Task" is described as "A task is a piece of work to be carried out within a project. There can be a number of tasks within the one project i.e. segmentation, translation, proofreading, desegmentation." (The Rosetta Foundation n.d.)
"Segmentation" is described as "When a project file has a large word count (i.e. anything over 5,000 words), the file has to be segmented. Segmentation appears on the project page the same way as translation and proofreading tasks. You claim the task, download the file and split it up into more manageable units for translation and proofreading. Once this is done, the segments are uploaded and ready to be translated. Once all segments have been translated and proofread, the files are merged back together in a process called desegmentation. The final output is one merged file to be downloaded by the organisation." (The Rosetta Foundation n.d.) and "Project Impact" as "The project impact section on the project details explains what the translation will be used for." (The Rosetta Foundation n.d.)

In addition to the FAQ section, there is a video page at http://trommons.org/static/videos/

It contains two videos for volunteers and two for organisations. Each of the videos is between 4 and 12 minutes long and contains an introduction to the UI elements of Trommons.

Once the volunteer has completed the task, they are able to upload the finished file (Image 10). The upload form contains links to the original source file, as well as latest version of the file and original task file. Alternatively, the volunteer is able to provide feedback.
**Image 10: Sample upload page**

**Finished processing?**

Upload your translated version of Montague.docx

Choose File  Montague.docx

Maximum file size is 2MB

Upload

**Cannot find the task file?**  Misplaced the original file or the latest uploaded file?

Click [here](#) to download the original project file.

Click [here](#) to download the original task file.

Click [here](#) to download the latest uploaded file.
3.2.4. Review and feedback, post-delivery

After uploading the finished project file, the volunteer is taken to the feedback screen (Image 11) which gives the opportunity of rating the file that the volunteer translated, or for the proofreader rating the translation after review. The form asks the volunteer to provide a "Star rating of 1-5, where 1 = 'Many Errors' and 5 = 'Few Errors'" in the categories for volunteers commenting the source file:

Corrections: Were there many mistakes in the source file?
Grammar: how was the use of grammar in the source file?
Spelling: Were there many spelling errors in the source file?
Consistency: Were there any errors in consistency in the source file?

By default, all categories have a 3-star rating already selected and users are able to change the star rating or leave them as they are. In addition, they are able to provide a comment.
For volunteers rating a translation after proofreading, the same screen appears with the exception of "translated" being replaced with "proofread".

Volunteers can then see the numeric star rating provided by the other user (volunteer or organisation), as well as the comments provided. However, once the task is completed, the organisation has the option of archiving the project. This also moves the task the volunteer has provided to the profile page, which results in the volunteer not being able to see the feedback anymore.

3.3. Trommons features and workflow comments

During the description of the Trommons platform, some aspects stood out as potentially either supporting or diminishing volunteer motivation, as well as either helping or hindering them in gaining experience in translation. While a full analysis based on user interface design principles would be beyond the scope of this paper, some comments will follow below in preparation for the main part of the analysis.

While the overall workflow of Trommons seems very suitable for volunteer localisation projects, there are some areas that might be problematic in regards to volunteer motivation, in particular the aspect of competence and gaining experience. The next chapters will first provide some thoughts on the aspects of profile and profile creation, task creation, search and accepting, working on tasks, getting help and task completion as well as review and feedback, and post-delivery.

3.3.1. Profile and profile creation

In addition to basic information like their language combinations, the volunteers also have the option of providing additional details in their biography section, which could be valuable for organisations in understanding the level of experience and specializations of the volunteer. However, that biography section is voluntary and a free form field without comments on which information could be relevant, which means that even if volunteers fill it in, it might not be useful information for the organisations. While organisations currently don't have the option of selecting which volunteer works on their tasks, having more information about the background of the person who claims a task could help them understand how much support that particular volunteer will need, and what information they might be missing. One way of achieving this could be by providing additional guidelines on filling in the biography section, which could include pointers on mentioning previous projects, years of experience, areas of personal interest that could provide information on subject matter expertise as well as employment history. While the information would still be voluntary, it might be beneficial for the volunteers to know which information helps organisations with supporting them in their task.

If in the future organisations are given the option of choosing and approaching volunteers they would like to work on their project, it would likely be useful to them if the biography information was at least
partially formalized by using forms that contain checkboxes and drop-down menus in order to allow for filtering of some of the basic parameters, in addition to further information provided in the current biography field.

The badges that are part of the volunteer profile currently contain a limited number of badges that are being awarded. While those badges are suitable for helping volunteers understand to which extent they have gone through the basic steps in the sign-up process, it could be expanded even further. To both, give volunteers the sense of progression and thereby the indication that they are increasing their competence which is being acknowledge and to give organisations information on how successful volunteers have been on Trommons in the past, badges could be awarded whenever volunteers complete a certain number of tasks, either tied to a certain feedback rating or just the delivery itself. Only counting deliveries would on one hand be very basic since it doesn't consider the question of how many tasks the volunteer previously accepted and didn't complete, or why the tasks weren't completed and therefore wouldn't provide information on areas the volunteer finds difficult as well as how reliably accepted tasks are also completed. It would also not provide information on how well the tasks are completed. On the other hand, tying the badges to additional parameters like the evaluation of quality could be problematic since as the analysis of the evaluation and feedback comments will show, the feedback is not homogeneous and therefore indirectly displaying the feedback on the volunteer's profile is potentially unfair and could diminish motivation even further than the feedback itself when not displayed on the profile.

The section of the profile displaying archived tasks is currently not showing the feedback the volunteer received. This means that if they don't see the feedback before the project is archived, they will not see the feedback at all. Since there are no guidelines for the organisations regarding how long they should wait before archiving a project, those organisations might do so immediately after the task has been delivered and feedback provided if they don’t know that the volunteers won't be able to see their comments or that of the proofreader.

In addition to being able to see the feedback at all, displaying it on the profile would allow the volunteer to see their progress by being given the option to view all the feedback on one screen side by side. This could give them a sense of whether they are in fact achieving their goal of increasing their experience and gaining competence, if that is a reason for them to volunteer. However, also for those who give other reasons, being able to read over comments might be motivating, especially if the comments are overall positive.

The archived tasks are also missing information on whether the task was completed on time and fully as well as the comments from the organisation or other volunteers.

The reference letter sent by the system on request only includes the tasks that the volunteer has worked on, not whether the task was completed well. As with the badges discussed previously, it could be both, beneficial for and detrimental to the motivation of volunteers if the element of quality was introduced in the reference letters. On one hand, if a volunteer is competent and receives positive feedback, having a letter that summarizes the feedback can be a stronger indicator of their competence to the person they intend to give the letter to. On the other hand, if the feedback is less positive due to issues or differing
opinions on what constitutes a task that was completed well between the volunteer completing the original task and the person providing the feedback, then the inclusion of the feedback in the reference letter could appear to be unfair.

3.3.2. Task creation, search and accepting

The option for volunteers to track individual organisations and their activities seems beneficial since it allows volunteers to see which tasks they could work on and thereby either increase their competence in a specific area by gaining more experience, or select tasks where they already feel confident in their competence in. However, volunteers are not able to see previous projects the organisation has completed, or tasks other volunteers are working on currently for that organisation or even the same project. This means that an opportunity for collaboration among volunteers of different languages or working on different parts of the project is lost, and that volunteers aren't able to learn from the experiences of other volunteers. If a translator was able to see the work of another translator, potentially without revealing the identity of that translator, and the feedback they received as well as changes that were made to that translation, they could improve the quality of their work for example by ensuring they use the same terminology. They could also feel more confident in their expertise by seeing whether other volunteers made the same decisions as them, or learn from the mistakes of other volunteers by seeing the corrections that were made to previous translations.

There are also some improvements in the information the organisations provide that could be made. While of the 145 organisations listed under http://trommons.org/org/search/ only 5 had no homepage listed and 21 no biography, the biographies varied in length from one line descriptions to more extensive ones. While it is possible that some of the organisations with no or only short biographies are not active, it might still be beneficial to provide guidelines for organisations detailing what kind of information is important for a translator or proofreader. It is also not clear how useful the links to the homepages of the organisations themselves are for volunteers. This could be subject of further analysis.

Organisations also have no space to provide volunteers with information regarding their projects outside the individual task. This means that they would have to host such information outside of Trommons and link to it in the task description. However, a spot-check that included 50 tasks showed that none included reference material on the task level. Two included reference material on the project level, which concerned technical details in regards to handling the source files, but no language specific information. On the project level, there was also one task that included a link to a website that was available in the target language of the task, the other 49 tasks included links to websites that were in languages in other than the task target language. The same could be observed on the organisation level.

None of the tasks included information on whether any translations had been done for that language or other languages previously, attachments that could be helpful for translation like previously completed translations or reference material like glossaries and style guides.

This suggests that there is a knowledge gap on the sides of the organisations, where the contact person requesting translations through Trommons either doesn't know what kind of information is important to
translators and proofreaders, or isn't aware of its importance or how to ensure that such information is available. Whether translators and proofreaders compile their own reference material, like glossaries, and send them to the organisations when they have completed their tasks is unclear but if they are, then it seems that this material would not necessarily reach other translators for related tasks.

In order to support organisations in developing and circulating information that helps translators, it might be beneficial to provide them with a platform like a wiki where they can compile such information, which is easily accessible for all volunteers working on the projects and which can also be edited easily. In addition to that, guidelines on what kind of information is useful for volunteers could be developed and provided to the organisations. Another option to consider could be the introduction of the volunteer consultant role on Trommons, who could support organisations in improving their information flow and other aspects of the localisation process.

Overall, the connection between tasks and projects isn't very clear. The main reason for that is that tasks and projects generally contain the same information and that projects don't group multiple languages. This means that for a volunteer there is only ever one task visible for a project which leads to superfluous duplication of some information and can lead to other information being missed, if the volunteer doesn't click on the project name on the task screen, and then again on the organisation link on the project screen. In addition it isn't clear where the proofreader can find comments the translator supplied or whether those are only visible to the organisation.

When volunteers accept a task, they are provided with a checklist that among other things asks them whether they will have enough time to work on the file and if they are capable of doing so. However, an inexperienced translator or proofreader might struggle to evaluate how long the task might take them. To support them in doing so and to learn how to estimate a task, some basic information like how many words a professional translator commonly can take on for a full working day and some points like whether they think they might need more time than experienced translator would and how much time they realistically can dedicate to the task. In order to help volunteers evaluate whether they are capable of working on the task, a checklist that might include questions like whether they have read the entire document or at least the key parts of it, whether they are familiar with the subject matter or whether they know where to find reference material to use for terminology or stylistic decisions might be useful.

Ideally, the volunteers would then be able to actually answer the questions in a task acceptance form and could decide whether they take on the task regardless of not feeling 100% competent. The specific points that they feel they are missing could then be addressed by the organisations, or by a volunteer mentor. A volunteer mentor for could be a volunteer who is an experienced translator and who can show how to research for terminology, how to deal with difficult phrases or what steps could be undertaken after the translation or proofreading process in order to ensure a higher quality output. A volunteer mentor could also be a subject matter expert, who answers questions on the topic. Mentoring could be added as a task like there are now translation, proofreading, desegmentation and segmentation. If a volunteer selects that they would like mentoring or that they require additional help with the task, a separate mentoring task can be created and part of the volunteer task stream, allowing other volunteers to accept the task and act as a mentor to that volunteer.

Additionally, the organisations could be made aware that there isn't enough reference material or that some parts of the text are difficult and might require extra checking. The benefit of being able to make
the organisations and other volunteers aware of such issues beyond the comment feature on tasks is that it would allow them to step in and prevent problems rather than having to deal with them later, either as a proofreader or as a contact person at the organisation. It would also mean that volunteer translators and proofreaders could be given the opportunity to improve their abilities and gain experience while working on a task, rather than hearing about shortcomings after the fact. While it might not be easy for everyone to be criticized while working on a task, it would nevertheless provide a safety net against feelings of incompetence and lowered motivation and could therefore lead to a higher rate of completed tasks after they were accepted, in addition to motivating volunteers to accept tasks again in the future. Additionally, it can provide them with a sense of progress when they first require mentoring with every project and then less and less.

The potential danger of such a system however is that a volunteer mentor might not always be suited to fill that role and that it is not necessarily easy to evaluate who is suited and to what extent there should be a selection process. Also, if volunteer translator/proofreader and volunteer mentor face difficulties with each other, this could be detrimental to the project and future motivation to accept further task, requiring mediation. Whether such mediation could be handled by the organisations or whether it would require additional volunteer roles, like community leader or volunteer mediator would have to be subject of further investigation.

A first step towards helping volunteers feeling competent and increasing their expertise would be to provide them with an idea of what a good translation could include and which steps to undertake in order to achieve a good quality level. In order to support the aspect of autonomy for motivation, the volunteer community could define the quality points and suggestions on how to achieve them, possibly in a body of knowledge or similar. However, some basic pointers like how to achieve terminology consistency, reminder to use a spell checker or reminder to ensure the translation is complete wouldn't require much preparation work but can help volunteers feel more confident in their work.

Another addition that could increase the motivation of volunteers to complete a task by increasing their feeling of competence might be adding information on where to find help directly to the screen that appears after they have accepted the task, as well as on the task overview. While in the next section some suggestions on information that could be included in the FAQ are listed, adding those suggestions also to the task screen and the task accepted screen could increase the visibility of the information and lower the number of volunteers who accept a task and didn’t complete it.

3.3.3. Working on tasks, getting help and task completion

While working on the task, volunteers are able to speak to the organisation who has requested it by using the comment field of the task. However, there is no immediately obvious option of interacting with the volunteers who are working on related tasks, for example those that are working on other languages or those that are providing a different part of the project, like other documents. Volunteer translators can leave comments for proofreaders in their documents, but proofreaders can’t discuss them with the translators. Also, since the project flow is such that the task for proofreading only gets created once the
translation is done, the proofreader will only claim the task after the translator has finished their part, which means that translator and proofreader won't be able to discuss their work while the translation is in progress. This could be especially challenging for the proofreader if a document is broken down into multiple elements as well as for the translator since they won't know what other translators are doing, or whether they are introducing inconsistencies between the different parts of the translation that the proofreader will have to deal with.

One way to avoid this situation might be to utilize the already existing discussion forums on http://forum.trommons.org/. However, the forum is very empty and there are only about a dozen threads, all of which only contain few replies. Only one thread contains information for the volunteers, the remaining ones are technical support requests. Whether this is due to lack of awareness that the forums exist, the desire to avoid posting to a near-empty forum or the attitude that discussing project specific questions on the forum isn't worthwhile would require additional investigation. While increased communication can't be forced and there may be volunteers who prefer to work alone without the social aspect, it could still be worthwhile to encourage organisations to start a discussion thread for their projects. They could then link to the forum in the description of their tasks, thereby allowing greater transparency and exchange for the volunteers via the forums. They could also write additional explanations in that forum thread and thereby share them with all volunteers at the same time.

Another option could be creating a communication area within Trommons that is automatically generated for each project and would close when the project is completed. The advantage of that solution would be that it avoids filling the forums with project specific discussions and also allows for greater privacy than a public conversation, thereby avoiding embarrassment for individual volunteers. It would also mean that organisations don't have to remember creating and linking to a discussion and that users don't have to create separate accounts for the discussion forums as is currently the case, which might deter them and lead to confusion due to mismatching user names, but could continue using their main Trommons login data. Another benefit is that users not directly involved in the project could voice their opinions and offer support.

However, the disadvantage would be a loss of knowledge since for future projects the volunteers would not be able to see what had been discussed on previous ones. Additionally, the forums might become disorganized due to a high volume of very project specific discussions. Another potential disadvantage is that while volunteers might benefit from users not directly involved in the project voicing their opinions or offering support, this might also create confusion if it isn't clear in what capacity users are posting on the project-specific thread.

Another element of feedback and support is that users might benefit from being able to retain the comments they exchanged with other users during the project. For organisations this could result in a knowledgebase they could compile and make accessible to volunteers, for the individual volunteer it could mean an overview of the issues they have encountered in the past and how they were solved, thereby giving them a sense of progression and the opportunity to review their learning points. This could be achieved by saving and displaying task specific comment threads on the volunteer and organisation profiles. It could also be achieved with a more customized approach by letting users tag specific comments or forum replies and thereby generating their personalized set of discussions they would like to retain for future reference.
On the FAQ pages, users can find some information on what some terms like tasks, segmentations and project impact are. This seems suitable to help with understanding the overall workflow. However, there are no further explanations regarding a task like segmentation and what best practices to follow. A volunteer might already know or intuitively understand how to segment text, but additional information on how to decide where to split the text, how big the chunks should be and considerations regarding formatting and data format could help with gaining knowledge and ensuring that the task is completed in a sensible way. Another example is the proofreading task. Since there is no definition of what is understood as proofreading, it might mean for some users that only objectively incorrect language is corrected, while for others it might mean adapting the entire translation to their preferences. Even basic questions like whether the proofreader should generally implement changes already, use track changes or just suggest changes in comments aren’t clarified. This could lead to confusion since on one hand proofreaders won’t know for sure what is expected of them while on the other hand if they interpret the proofreading task to mean that they should adapt the translations to their preferences even if the translation is objectively correct, it could lead to an increased number of changes and harsher criticism for the translator. This in return could lower their motivation since it implies they are less competent than they are in reality unless they understand that the changes made were preferential. While not every project is the same and some come with differing requirements, a basic working definition could serve as the foundation for further discussion and clarification of project specific tasks.

Fields like project impact, task description and reference require further clarification for the organisations. As described, a spot check of 50 tasks showed very little information for the translators and proofreaders that would support their work. The FAQ page might be a suitable platform for clarifying this topic, or alternatively a wiki page where organisations and volunteers can interact with each other and less experienced contact persons on the organisations side could learn from those that are more experienced with translation projects.

While the FAQ page encourages volunteers to contact The Rosetta Foundation if they require support, an addition to this could be to first list other options like checking the forums for help, watching the support videos, contacting the organisation who has requested the task as well as checking the task, project and organisation descriptions for reference material. Some questions could also be answered by other volunteers, in particular those who have more experience with Trommons projects.

While it is positive that The Rosetta Foundation is available for help, suggesting additional alternatives for support could increase the autonomy of Trommons volunteers and also show organisations what the needs of volunteers are. By creating a wider support net for low level issues, volunteers could also get support at times where The Rosetta Foundation is not staffed, for example due to differing time zones. This could mean that a volunteer who is currently stuck on a project might be able to continue working again on the same day, rather than the next, especially if the issue is a minor one.

Another venue for users on Trommons to get information are the support videos explaining the user interface and different features. While these videos seem suited in increasing the competent use of the user interface, they are missing an explanation of what they contain and why the user should watch them. Since they are relatively long, it seems likely that not a lot of users watch the videos fully and are there for missing information which could help them increase their competence in using the Trommons platform. In addition to a description, searchable transcripts with time stamps could also help users find
the information that is important to them quickly and thereby help them utilize the existing features on Trommons correctly and fully.

The original direction of this paper included investigating the number of volunteers who accepted a task but didn’t complete it, their distribution within the user base and the question whether a loss of motivation was to be blamed for the decision to first accept a task, which indicates that they were initially motivated to work on it, and then not complete it, which may indicate that an event or circumstance related to the task or Trommons may have lowered their motivation. Unfortunately, in the course of the research process it became evident that too little data on users who previously accepted a task and then didn’t deliver it was being logged, which means that a change of direction was necessary. This situation has now been corrected, which means that after a few months of data logging further investigations into this question should be possible.

However, according to anecdotal evidence provided by Alison Nolan, Production Coordinator at The Rosetta Foundation, and Paulina Abzieher, Community Coordinator at The Rosetta Foundation, volunteers abandon tasks on a regular basis without unclaiming them. In this case, they usually receive two e-mails from TRF staff and if they don’t react to the reminders, the task is unclaimed for them and a different volunteer found to take on the task. While there is no concrete data at this point, this may suggest that volunteers require more information on how to unclaim tasks and what to do if they are stuck or how to evaluate their competence level and time requirements for the task.

This might be especially important since it seems reasonable to assume that at least some volunteers don’t accept further tasks after having accepted and then abandoned one previously, since if a lack of motivation is the reason for this, that motivation could be low also in the future. This would result in a certain number of previously motivated, but now inactive volunteers on Trommons.

3.3.4. Review and feedback, post-delivery

In addition to the volunteers not being able to see their feedback after a task has been archived as discussed previously, there are also some areas in the information flow that could be improved regarding feedback and evaluation of tasks.

In the FAQ, organisations are provided with information on how to give feedback to volunteers from a technical perspective, which should increase their confidence in navigating the user interface competently. However, there is no information on what constitutes good feedback, what helps volunteers improve and how to evaluate whether a task was done well. While it is unclear to what extent the organisations themselves are able to evaluate the quality of a delivery, at least from a linguistic standpoint, they likely can evaluate process related aspects to some extent. In addition, they could gather information on how suitable the translation was for its purpose and for its target audience from the target audience directly after publishing in case they don’t have a separate evaluation process. This target audience feedback could then provide valuable information for the volunteers and help them increase their feeling of competence and improve their skills. However, in order to motivate organisations to gather such information and provide it to the volunteers, some explanations on what to do, how and why would likely be needed.
As the following analysis of the numeric star rating and feedback comments will show, the lack of guidelines for organisations as well as proofreaders in regards to how to provide feedback and rate a task that allows translators to improve their skills and be motivated to accept further tasks has a direct impact on the quality of the feedback and numeric ratings.

Another factor to consider is that the feedback happens after the task has been completed, and is unilateral without giving translators the opportunity to explain their choices or asking additional questions. This gives the feedback more the character of evaluation than of feedback intended to support the receiver in improving their skills.

4: Output review and feedback

4.1. Composition of volunteer data

Within Trommons.org, volunteers and organisations have the opportunity to rate and comment on the work provided. The categories the volunteers' work can be rated in are Corrections, Grammar, Spelling and Consistency. A sheet containing task lines with the associated Task ID, numeric star rating per category, comment field and user name was made available by TRF for the purpose of this research.

Since the actual nature of the errors that were detected was not the focus of this investigation, the average of those four was calculated and used for analysis in order to cut down on variables that needed to be considered. This means that while Trommons only allows whole numbers, the numbers for analysis also included the fractions 0.25, 0.5 and 0.75.

After an initial examination of the data, some observations could be made:

Not all users who had provided a numeric star rating had also left a comment. This was not only the case for those lines where all categories had been rated as 3 stars, which is the default setting in Trommons, but also for other numeric star rating categories.

Not all users seemed to have understood the way the rating should be done correctly. This was observed by comparing the star rating with the comment. For example, if the comment was very positive but the rating low, it could be considered a rating where the user had not understood the rating system correctly. Others were ambiguous and placed in their own category.

Some users were present in very high numbers, in multiple tasks while others appeared only in one or two lines. While this is not necessarily a concern for the analysis, it could nonetheless be relevant when considering whether the data might be distorted by the way individuals provide feedback.

The data sheet containing reviews of task outputs (deliveries) contained a total of 658 lines. In order to address the observations outlined above, the initial data processing steps were:
1 - Split the lines without comments from those with comments. The assumption is that based on the comments, it is possible to evaluate whether the point rating system was used correctly and intentionally. Therefore, no evaluation is possible whenever no comments were provided.

2 - Count unique users in each categories.

3 - Compare users of each categories and determine whether any of them are present in both.

4 - Count how often each user that is present in both occurs in the respective category.

The result of these steps were as follows:

A total of 325 lines, split across 167 users, had no comment associated with them. 333 lines, split across 129 users, included comments.

48 users were in both categories, those that had provided comments and those that hadn't.

6 users had left 5 or more comments, with the highest being 27 and 8 - both of which are TRF employees. The first volunteer in the list left 7 comments. 20 users left between 2 and 4 comments, while 23 users provided only 1 comment.

8 users had submitted a high amount of work without commenting, the highest being 11 and 10 submissions, both of which were volunteers. 20 users submitted between 2 and 4 tasks without comment, and 20 users submitted 1 task without commenting.

This means that there is no significant difference in the distribution of users who leave comments and those who don't.

18 users submitted an equal amount of tasks with and without comment, of which 4 users submitted 2 tasks in either category and 14 users 1 task in each.

18 users submitted more tasks without comment than with, 12 submitted more with comments than without.

This means that a nearly equal number of lines had no comments, and nearly a third of users sometimes left comments and sometimes not.

After the initial analysis of user distribution, the lines were categorized as follows:

5 - The comments were compared with the numeric rating to determine whether the user had likely understood the rating system correctly. The categories given were:
Yes: The comments and numeric rating clearly match. For example, the user commented that the translation was very good and also provided a very high rating, or the user commented that the translation was poor and provided a low rating.

No: The comments and numeric rating clearly did not match. For example, the user commented that the translation was very good but provided a low rating. Included in this category were users who had provided numeric rating in the comments rather than using the radio button option.

Unsure: It is not clear from the comments whether the numeric rating was chosen intentionally to reflect a quality evaluation. Commonly, this included such comments that didn't comment on the quality at all, so no clear evaluation was possible. It also included those tasks where what seemed to be a contact person from the organisation had thanked the volunteer for their work or had used the comment to convey messages other than quality evaluations.

Additionally, numeric ratings of 3 in all categories, which is the default setting in Trommons, were placed in the category unsure/unchanged, since it is unclear whether the user chose to not alter the rating for any of the categories to express an opinion on the quality of the delivery, or whether they left the default without further consideration.

The result of these steps were:

**NULL Category**

Of the 325 lines where no comment was provided ("null" category), 88 had received an average rating of 5, which means that the user had chosen the highest rating for all areas. 88 ratings were between 4.25 and 4.75, which means that the user had chosen ratings between 3 and 5 for the different areas, and had chosen different numbers for the different areas. 35 had received an average rating of 4, of which 18 had been rated 4 in all areas.

41 tasks in the "null" category had received average ratings between 3.25 and 3.75 with the individual values for the areas ranging from 1 to 5, sometimes within the same task.

43 tasks in the "null" category had received average ratings of 3, of which 32 had 3 assigned to all areas.

30 tasks received an average rating between 1.25 and 2.75, with 3 tasks receiving a rating of 2 in all areas, and none a rating of 1 in all areas.

This means that approximately half of the lines without a comment were rated high or very high, which could indicate that users who do not wish to spend time on writing an evaluation in comment form, still want to express their positive view of the task. It is especially interesting that half of those rated high or very high were 5 star ratings, which means that 5 stars had been selected across the board. By comparison, the number of lines where the default 3 star rating for all lines remained is remarkably low in this category, considering that giving an all-5 rating requires more effort than an all-3 rating, albeit only slightly. This could indicate that the user chose the highest possible number in order to express appreciation and encouragement, without actually wanting to make a statement on the quality.
However, since there were no comments associated with these lines, it was not possible to confirm this hypothesis.

**YES Category**

Of the 165 lines in the "yes" category, 60 received an average rating of 5, which means that all areas received top marks. However, in only 32 cases, the comments expressed that no corrections were made or that the translation was excellent or perfect, thereby justifying the full rating in all categories. 28 comments mentioned slight corrections that were made, ranging from formatting changes to wording and spelling corrections. The 32 entirely positive comments were split among 19 users, the 28 containing corrections also were provided by 19 individual users. 4 users appear in both categories.

53 tasks in the "Yes" category had received ratings between 4.75 and 4.25, split among 44 users. It is worth noting that the 21 tasks in this category with a rating of 4.75, had comments that were similar to those with a rating of 5 across all areas with slight corrections being in the comments. Reviewers of both categories describe translations of great, perfect or good and then mention that they made few corrections or changes. 21 tasks were rated as 4.5, and the comments were also mostly positive, 11 were rated 4.25.

14 tasks, rated by 14 users, received an average rating of 4, of which 5 had 4 stars in all categories, while 10 had ratings between 1 and 5 in individual categories.

17 tasks were rated between 3.75 and 3.25, and the ratings were done by 13 individual users.

6 tasks, rated by 6 individual users, received an average rating of 3. 2 tasks had ratings of 3 in all categories.

13 tasks, rated by 16 users, received an average rating of 2.75 to 1.25. 1 task received a rating of 2 in all areas.

2 tasks in the "yes" category received a rating of 1 in all areas, and the comments were very negative.

This means that more than half of the ratings that had a comment associated with them were placed in the "yes" category. Around two-third of the "yes" lines were rated high or very high.

**NO Category**

For 22 tasks, split across 12 users, the comments showed that the person evaluating and commenting had not understood how to use the rating system correctly. This could be seen in two ways:

15 tasks had numeric rating in the comments. For all of these tasks, the rating had been left as 3 in all areas, indicating that the user had misunderstood how to use the rating system. There were 6 users who had provided feedback in this format.
For 7 tasks, the comments were much more positive than the numeric rating. For those 7 tasks, the commenter had indicated that they had made no changes or that the translation was very good, requiring only few changes. However, for 2 of the tasks they had left the numeric rating as 3 in all areas, and for 5 had rated it as 1 or 2 in all areas, indicating that the user had misunderstood which rating indicated "good" and which "not good".

7 users in the "No" category had also commented on tasks that were in other categories. Most of the tasks were in "Null" and "Unsure", but some also in the "Yes" category.

This shows that only a very small number of users clearly didn't understand how the rating system works, which indicates that it is successful in being simple and intuitive to use.

**UNSURE Category**

The "Unsure" category contained tasks where it was either unclear whether the user had understood the rating system correctly or where no comments regarding the quality were made, which means it wasn't possible to evaluate whether the numeric rating had been given intentionally as a statement on quality. This category also included most ratings where the default value 3 hadn't been changed in any of the areas except for those where the comment indicated that this may have been intentional.

The category "Unsure" consisted of 146 tasks, split across 55 users.

51 tasks, split across 15 users had a rating of 5, which means that 5 had been given for all areas. Of those 51 tasks, 48 split across 12 users contained no quality comment, but instead usually a comment on the nature of the task or in most cases thanked the volunteer for their contribution without providing feedback on the task. 2 tasks seemed like they may have been rated too high, while 1 was in a language other than English and could therefore not be evaluated.

5 tasks had received numeric ratings between 4.75 and 4.25, none of which contained a quality comment.

None had a rating of 4, 1 task was rated in the category between 3.75 and 3.25 and could be considered a low rating.

78 tasks had a rating of 3, split across 27 users. 59 of those contained no quality comment, all but one of those 59 had a rating of 3 across all areas. 18 of the 78 tasks had a quality comment that was either negative or positive, but since all areas had been left with the default rating of 3, they were placed into the "unsure" category.

The category from 2.75 to 1.25 contained 5 tasks, of which 1 could be considered a low rating and 4 contained no quality comment.

1 task received a rating of 1 in all areas, but the comment was unclear so it wasn't obvious whether the rating had been intentional.

The remaining task lines were incomplete and had to be ignored.
Overall, while the rating system appears to be clear and intuitive enough to be used correctly by most users, the comment field is underutilized. Only about a quarter of the lines contained comments that could be considered for further analysis, while the remaining 3 quarter either contained no comment or comments that were not suited for analysis. This suggests that only for about a quarter of the tasks included in the data sheet, the volunteers had received feedback beyond the numeric rating that was a comment on the quality without being confusing due to conflicting information or a statement not related to a quality remark.

4.2. Coding and analysis of the comments

4.2.1. Coding and analysis method

For the analysis of the comments, only those feedback lines that were categorized as "YES", thereby indicating that by comparing the comment with the star rating it was reasonable to assume that the user had understood the rating system and was using it as intended, were used. The rationale behind this was that in order to be able to compare the numeric star rating with the sentiment in the comments, those where it wasn't clear that the star rating system had been used as intended had to be excluded since otherwise noise from users misunderstanding the system or ambiguous ratings might have distorted the picture. In order to be able to come to conclusions also for the following parts of the analysis, all but the "YES" comment lines were excluded also for them.

As a result, 165 comments were coded using the magnitude coding method as described by Saldana (2009, pp. 72-7) in order to be able to gather information on the frequency and distribution of the different types of comments. The main research question was broken into sub-questions, based on which the following codes were used in the coding process:

1 - What is the distribution of positive, neutral and negative remarks on the work of the volunteers, relative to the numeric rating given to the task? (Table 1)

<table>
<thead>
<tr>
<th>Code name</th>
<th>Explanation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOOD</td>
<td>Positive remark</td>
<td>Good, understandable, consistent</td>
</tr>
<tr>
<td>VGOOD</td>
<td>Very positive remark</td>
<td>Perfect, great, nothing to change</td>
</tr>
<tr>
<td>BAD</td>
<td>Negative remark, highlighting issues</td>
<td>Problems, inappropriate, typing errors, poor, literal, not fluent, mistake, corrected, confusing, not translated, awkward</td>
</tr>
</tbody>
</table>
2 - Are the remarks worded in a way that is suitable to encourage volunteers rather than discourage them? (Table 2)

<table>
<thead>
<tr>
<th>Code name</th>
<th>Explanation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Increasing term</td>
<td>Very, totally</td>
</tr>
<tr>
<td>0</td>
<td>Weakening term</td>
<td>Few, overall, just, some, quite, pretty much, a few, minor, almost, in general, in my view</td>
</tr>
<tr>
<td>ATT</td>
<td>Attack on volunteer, making assumptions</td>
<td>Stating the resource was from a country other than what was required, claiming google translate had been used, claiming translation hadn’t been reviewed prior to delivery</td>
</tr>
</tbody>
</table>

It is important to note that comparing the feedback and comments to the actual work that had been delivered was not within scope of this paper. This means, that it is conceivable that what appears to be a negative remark is objectively not as negative as it could have been if all the tasks had been evaluated by the same person and against the same set of criteria. However, while this aspect is relevant and could provide the foundation of further research, it is still useful to conduct an analysis of the comments in relation to the numeric rating since as discussed previously, perceived competence and feedback have an impact on motivation.

The codes + and 0 were combined with GOOD, BAD, VGOOD and VBAD to form codes like 0BAD or +BAD. This results in the sub-codes shown in Table 3. To answer sub-question 1 "What is the distribution of positive, neutral and negative remarks on the work of the volunteers, relative to the numeric rating given to the task?", the sub codes carrying increasing or neutralizing/weakening adjectives were categorized as either GOOD, VGOOD, BAD, VBAD since the only differentiating aspect is that they carry an additional adjective, while the remark itself makes it suitable to be categorized as either of those main codes.
<table>
<thead>
<tr>
<th>Subcode</th>
<th>Explanation</th>
<th>Examples</th>
<th>Categorized as</th>
</tr>
</thead>
<tbody>
<tr>
<td>+GOOD</td>
<td>Increasing term combined with positive remark</td>
<td>Very good, totally understandable</td>
<td>VGOOD</td>
</tr>
<tr>
<td>0+GOOD</td>
<td>Neutralizing/weakening term combined with increasing term and positive remark</td>
<td>Overall very well translated, mostly very good</td>
<td>VGOOD</td>
</tr>
<tr>
<td>+VGGOOD</td>
<td>Increasing term combined with very positive remark</td>
<td>Very impressive</td>
<td>VGOOD</td>
</tr>
<tr>
<td>0VGGOOD</td>
<td>Neutralizing/weakening term combined with very positive remark</td>
<td>Almost perfect</td>
<td>VGOOD</td>
</tr>
<tr>
<td>00VGGOOD</td>
<td>Two neutralizing/weakening terms combined with very positive remark</td>
<td>Overall almost perfect</td>
<td>VGOOD</td>
</tr>
<tr>
<td>0GOOD</td>
<td>Neutralizing/weakening term combined with positive remark</td>
<td>Overall, well done</td>
<td>GOOD</td>
</tr>
<tr>
<td>00GOOD</td>
<td>Two neutralizing/weakening terms combined with positive remark</td>
<td>In general fairly decent</td>
<td>GOOD</td>
</tr>
<tr>
<td>+BAD</td>
<td>Increasing term combined with negative remark</td>
<td>Multiple errors, many typos</td>
<td>VBAD</td>
</tr>
<tr>
<td>++BAD</td>
<td>Two increasing terms combined with negative remark</td>
<td>Much too literal</td>
<td>VBAD</td>
</tr>
<tr>
<td>0+BAD</td>
<td>Neutralizing/weakening term combined with increasing term and negative remark</td>
<td>At times very literally</td>
<td>VBAD</td>
</tr>
</tbody>
</table>
Neutralizing/weakening term combined with negative remark

Sometimes incorrect, some corrections

BAD

Two neutralizing/weakening terms combined with negative remark

Just a few improvements, just a few problems

BAD

Three neutralizing/weakening terms combined with negative remark

Only a few minor corrections, perhaps a little too literally at times

BAD

It should be noted that expressions coded as +GOOD could have been combined into the main code VGOOD. However, in order to avoid bias during the coding process, it was useful to only assign VGOOD when the term itself was a very positive remark like "no changes" or "excellent" and assign +GOOD whenever the remark could have stood by itself without the increasing term, like "very good" or "totally understandable". The same applies for combinations of + and BAD.

3 - Are the remarks suited to help volunteers improve their work? (Table 4)

Table 4

<table>
<thead>
<tr>
<th>Code name</th>
<th>Explanation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>HELP</td>
<td>Remark helps volunteers to improve their work by pointing out specific issues or giving concrete improvement suggestions, remark indicates which specific areas the volunteer already did well in</td>
<td>Pointing out spelling, grammar, terminology etc.</td>
</tr>
</tbody>
</table>

A single remark may, and was commonly, coded both with HELP and one of the codes or sub-codes indicating whether the remark is positive or negative. For example, if a comment said that the translation was "very good", that remark only received the code +GOOD. If the comment said that the spelling was "very good", the remark received both, the codes +GOOD and HELP.
4 - Additional codes (Table 5)

Additionally to or instead of evaluating the quality of the translations, some volunteers or organisations also remarked on issues they encountered, or whether they liked or disliked the task of proofreading the text. While issues or the satisfaction of volunteers in the proofreading step is not within scope of this paper, such occurrences were coded regardless in order to be able to completely cover all types of comments.

Table 5

<table>
<thead>
<tr>
<th>Code name</th>
<th>Explanation</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISSUE</td>
<td>Comment on issues regarding the task overall, not related to the work provided by the volunteer</td>
<td>Source file leaves to be desired, lacked context</td>
</tr>
<tr>
<td>LIKE</td>
<td>Comment expressing pleasure of doing the job</td>
<td>Pleasant, pleasure, enjoyed</td>
</tr>
<tr>
<td>DISLIKE</td>
<td>Comment expressing displeasure of doing the job</td>
<td>Took much longer than thought, took a long time</td>
</tr>
</tbody>
</table>

As a first analysis step, the number of occurrences for each type of remark was counted for each comment. Then the comments were sorted by average numeric rating associated with them and placed into categories from 5.0 to 1.0 in 0.25 star increments. Then the occurrences for each type of remark were added up for each category of numeric rating. (Table 6)
<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0+GOOD</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>+GOOD</td>
<td>20</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VGGOOD</td>
<td>30</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0VGOOD</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>00VGOOD</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>+VGGOOD</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GOOD</td>
<td>19</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0GOOD</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>00GOOD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NEUT</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BAD</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>13</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>0BAD</td>
<td>14</td>
<td>5</td>
<td>9</td>
<td>7</td>
<td>9</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>00BAD</td>
<td>16</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>000BAD</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>+BAD</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0+BAD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>++BAD</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>VBAD</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FAY</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>HELP</td>
<td>19</td>
<td>10</td>
<td>21</td>
<td>15</td>
<td>15</td>
<td>10</td>
<td>2</td>
<td>13</td>
<td>12</td>
<td>9</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>12</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>LIKE</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>DISLIKE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ISSUE</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Since the number of tasks within each category varied greatly, the numbers of occurrences of a remark in each category were divided by the overall number of comments linked to a rating average category. For example, there were 60 comments linked to the 5 star rating average category. Therefore, the number of occurrences of the code 0+GOOD were divided by 60, resulting in the number 0.033, rounded. Using this method, it is possible to show the number of occurrences of a remark per comment without distorting the numbers due to the fact that the some categories of numeric rating were much more prevalent than others, and thereby received a higher number of comments and by extension a higher number of individual remarks. (Table 7)

These equalized numbers were then used as the basis of further analysis to answer the sub-research questions.
Table 7

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nr. of comments</td>
<td>80</td>
<td>90</td>
<td>95</td>
<td>100</td>
<td>110</td>
<td>120</td>
<td>130</td>
<td>140</td>
<td>150</td>
<td>160</td>
<td>170</td>
<td>180</td>
<td>190</td>
<td>200</td>
<td>210</td>
<td>220</td>
<td>230</td>
</tr>
<tr>
<td>0GOOD</td>
<td>0.031</td>
<td>0.030</td>
<td>0.029</td>
<td>0.028</td>
<td>0.027</td>
<td>0.026</td>
<td>0.025</td>
<td>0.024</td>
<td>0.023</td>
<td>0.022</td>
<td>0.021</td>
<td>0.020</td>
<td>0.019</td>
<td>0.018</td>
<td>0.017</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td>0.5GOOD</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td>0.186</td>
<td></td>
</tr>
<tr>
<td>1GOOD</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td>0.400</td>
<td></td>
</tr>
<tr>
<td>1.5GOOD</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td>0.017</td>
<td></td>
</tr>
<tr>
<td>2GOOD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>2.5GOOD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>3GOOD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>3.5GOOD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>4GOOD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>4.5GOOD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>5GOOD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>BAD</td>
<td>0.017</td>
<td>0.125</td>
<td>0.203</td>
<td>0.293</td>
<td>0.393</td>
<td>0.500</td>
<td>0.600</td>
<td>0.727</td>
<td>0.843</td>
<td>0.970</td>
<td>1.097</td>
<td>1.224</td>
<td>1.351</td>
<td>1.478</td>
<td>1.605</td>
<td>1.732</td>
<td></td>
</tr>
<tr>
<td>BAD</td>
<td>0.267</td>
<td>0.376</td>
<td>0.485</td>
<td>0.594</td>
<td>0.703</td>
<td>0.812</td>
<td>0.921</td>
<td>1.030</td>
<td>1.139</td>
<td>1.248</td>
<td>1.357</td>
<td>1.466</td>
<td>1.575</td>
<td>1.684</td>
<td>1.793</td>
<td>1.902</td>
<td></td>
</tr>
<tr>
<td>BAD</td>
<td>0.017</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>HABD</td>
<td>0.017</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>HBAD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>VBAD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>ATT</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>HELP</td>
<td>0.117</td>
<td>0.256</td>
<td>1.500</td>
<td>1.364</td>
<td>1.500</td>
<td>1.250</td>
<td>1.250</td>
<td>1.250</td>
<td>1.250</td>
<td>1.250</td>
<td>1.250</td>
<td>1.250</td>
<td>1.250</td>
<td>1.250</td>
<td>1.250</td>
<td>1.250</td>
<td></td>
</tr>
<tr>
<td>LIKE</td>
<td>0.017</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>DISLIKE</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>ISSUE</td>
<td>0.000</td>
<td>0.100</td>
<td>0.100</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>
4.2.2. Analysis results

4.2.2.1. What is the distribution of positive, neutral and negative remarks on the work of the volunteers, relative to the numeric rating given to the task?

In order to show the distribution of positive, neutral and negative remarks relative to the numeric rating given to the task the equalized numbers were added into the groups "Very positive", "Positive", "Neutral", "Negative", "Very negative" (Table 8) and their values added up (Table 9, Image 12).

*Table 8 - Groups with codes and subcodes*

<table>
<thead>
<tr>
<th>Group</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very positive</td>
<td>0+GOOD</td>
</tr>
<tr>
<td>Very positive</td>
<td>+GOOD</td>
</tr>
<tr>
<td>Very positive</td>
<td>VGOOD</td>
</tr>
<tr>
<td>Very positive</td>
<td>0VGOOD</td>
</tr>
<tr>
<td>Very positive</td>
<td>00VGOOD</td>
</tr>
<tr>
<td>Very positive</td>
<td>+VGOOD</td>
</tr>
<tr>
<td>Positive</td>
<td>GOOD</td>
</tr>
<tr>
<td>Positive</td>
<td>0GOOD</td>
</tr>
<tr>
<td>Positive</td>
<td>00GOOD</td>
</tr>
<tr>
<td>Neutral</td>
<td>NEUT</td>
</tr>
<tr>
<td>Negative</td>
<td>BAD</td>
</tr>
<tr>
<td>Negative</td>
<td>0BAD</td>
</tr>
<tr>
<td>Negative</td>
<td>00BAD</td>
</tr>
<tr>
<td>Negative</td>
<td>000BAD</td>
</tr>
<tr>
<td>Very negative</td>
<td>+BAD</td>
</tr>
<tr>
<td>Very negative</td>
<td>0+BAD</td>
</tr>
<tr>
<td>Very negative</td>
<td>++BAD</td>
</tr>
<tr>
<td>Very negative</td>
<td>VBAD</td>
</tr>
</tbody>
</table>
Table 9 – Equalized numbers in groups

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very positive</td>
<td>0.850</td>
<td>0.579</td>
<td>0.350</td>
<td>0.273</td>
<td>0.357</td>
<td>0.000</td>
<td>0.000</td>
<td>0.250</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.250</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Positive</td>
<td>1.133</td>
<td>0.789</td>
<td>0.800</td>
<td>0.636</td>
<td>0.714</td>
<td>0.375</td>
<td>0.000</td>
<td>0.750</td>
<td>0.500</td>
<td>0.000</td>
<td>0.500</td>
<td>1.000</td>
<td>0.250</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.017</td>
<td>0.000</td>
<td>0.000</td>
<td>0.091</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Negative</td>
<td>0.533</td>
<td>0.895</td>
<td>1.500</td>
<td>1.364</td>
<td>1.214</td>
<td>1.250</td>
<td>1.250</td>
<td>0.833</td>
<td>3.000</td>
<td>2.500</td>
<td>1.000</td>
<td>0.750</td>
<td>3.333</td>
<td>3.000</td>
<td>0.000</td>
<td>3.000</td>
<td></td>
</tr>
<tr>
<td>Very negative</td>
<td>0.017</td>
<td>0.000</td>
<td>0.100</td>
<td>0.091</td>
<td>0.000</td>
<td>0.250</td>
<td>0.500</td>
<td>0.750</td>
<td>0.167</td>
<td>0.750</td>
<td>0.000</td>
<td>1.500</td>
<td>1.000</td>
<td>1.333</td>
<td>3.000</td>
<td>1.000</td>
<td>2.000</td>
</tr>
</tbody>
</table>

Image 12 – Visual representation of equalized numbers in groups
While the number of comments analysed is higher in the numeric rating groups 3.75 and above than in the groups below, which means that for some of the spikes in the lower groups the data is insufficient to draw reliable conclusions like the spike in negative remarks at 2.75 which is higher than the occurrences in most lower categories, some observations can be made:

The overall trend of the comments reflects the numeric star rating insofar that the top 5 groups 5 to 4 combined contain significantly more positive and very positive remarks per comment than the middle 6 3.75 to 2.5 and the bottom 6 2.25 to 1. Likewise, the top 5 groups contain fewer negative and very negative remarks per comment than the middle 6 and bottom 6 (Table 10).

<table>
<thead>
<tr>
<th></th>
<th>5 to 4</th>
<th>3.75 to 2.5</th>
<th>2.25 to 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Very positive</strong></td>
<td>1.296</td>
<td>0.396</td>
<td>0.250</td>
</tr>
<tr>
<td><strong>positive</strong></td>
<td>1.143</td>
<td>2.125</td>
<td>3.486</td>
</tr>
</tbody>
</table>

However, when looking at the detailed numbers (Table 9), the numeric ratings don’t correlate with the comments insofar that the trend of decreasing numbers of very positive and positive remarks shows irregularities. The group 4.5 and 4 have nearly the same number of very positive remarks (0.350 and 0.357 respectively) while 4.25 is lower than both of them with 0.273 very positive remarks per comment. 3.75 and 3.5 include no very positive remarks while 3.25 has nearly as many positive remarks per comment as 4.25 (0.273). While 3 to 2.25 as well as 1.75 to 1 include no very positive remarks, 2 includes as many per comment as 4.25 (0.025). Even when ignoring the groups below 3.75, this suggests that there is no significant difference between a 4.5 and a 4.0 rating. Overall, the number of very positive remarks per comment also doesn’t show a clear trend and remains relatively flat. There are multiple numeric rating categories that have no very positive remarks, and for those that do, the lowest number is 0.250 and the highest 0.850. The curve is even flatter when looking only at the categories 4.5 and lower, where the highest number of very positive remarks per comment is 0.375.

Similar observations can be made for the numbers of positive remarks per comment. The categories 4.75, 4.5 and 4.0 have a very similar number of positive remarks per comment (0.789, 0.800 and 0.714 respectively) while 4.25 with 0.636 is lower than 4.0 which shows no significant difference between an average 4.75, 4.5, 4.25 or 4.0 star rating. When looking at the rating averages below 3.75, there are no positive remarks in the 3.5 category, while 3.25 contains with 0.750 positive remarks per comment nearly as many as the 4.75 category. 3 and 2.5 contain with 0.500 positive remarks per comment more
than the 3.75 category and 2.25 with 1.000 more than any other category in the positive remark group with the exception of the 5 star rating average, which has 1.133 positive remarks per comment. On the other hand, 3.75 and 2.75 contain no positive remarks.

The overall curve for positive remarks is similarly flat as that for very positive remarks. Here there are also multiple numeric rating categories that have with 0.000 remarks, and for those that do have positive remarks, the lowest is 0.250 and the highest 1.133. While this is four times as many positive remarks for the highest as for the lowest, the numbers are spread out along the curve with only a slight downward trend from 5 to 4, with multiple subsequent spikes, especially around 3.25 and 2.25.

In the negative and very negative remarks categories, it should be noted that there are overall 3 to 4 times more negative and very negative remarks per comment than positive or very positive. This is especially remarkable since the star ratings show the opposite trend. (Table 11)

Table 11 - Number of comments per star rating group

<table>
<thead>
<tr>
<th>Number of comments</th>
<th>5 to 4</th>
<th>3.75 to 2.5</th>
<th>2.25 to 1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>124</td>
<td>28</td>
<td>13</td>
</tr>
</tbody>
</table>

Another factor is that the categories below 2.0 (1.75 to 1) contain no positive or very positive remarks, while the negative and very negative remarks are still present in the highest 5 to 4 star ratings).

This leads to the conclusion that users tend to be more detailed and/or repetitive when talking about negative aspects of a task they evaluate than when describing positive ones. This could also be observed during the coding process, where there was a strong tendency for users to list different aspects that needed improvement like grammar, spelling, terminology, formatting or style while they were much more general when discussing positive aspects.

It also shows that the majority of tasks (124) included in the analysis were rated very high or high (5 or 4 star average) while only 41 were high medium to low (3.75 to 1 star average). While this does not necessarily allow for conclusions regarding the objective quality of the work provided, whichever way that objective quality may be defined, it does show that users tend to see the tasks of others in a positive light, and wish to provide them with positive feedback.

When further analysing the tendencies in the negative and very negative remarks categories, similar trends to those in the positive and very positive categories can be observed.

Within the negative remarks group, the 5 and 4.75 star rating averages have 0.533 and 0.895 negative remarks per comment, which is similar to the 3 star rating (0.833). In between, the 4.25 to 3.25 star rating groups are remarkably consistent with 1.500 negative remarks per comment as the highest number and 1.250 as the lowest. These numbers double in the 2.75 to 1.0 star rating categories with 3.000 as the highest and 2.500 as the lowest number of negative remarks per comment, only the 2.25
(1.000), 2.0 (0.750) and 1.25 (0.000) star ratings deviate from this trend which might be explained by the overall low number of comments in those categories.

There are very few negative remarks per comment in the star ratings categories 5 to 4, with 0.000 being the lowest and 0.100 the highest. Between 3.75 and 2.75 that number climbs, making the 3 star rating average the lowest with 0.167 very negative remarks per comment and 3.25 and 2.75 the highest with 0.750 very negative remarks per comment. 2.5 (0.000) breaks the trend, followed by 2.25 to 1 with 2 or even 3 times as many very negative remarks per comment as the 3.75 to 2.75 categories (3.000 highest and 1.000 lowest).

Despite the lower number of comments in those low star rating categories, there is a clearer correlation of negative or very negative remarks per comment than for the positive or very positive ones. This suggests that while users overall tend towards giving high numeric ratings, they don’t hesitate to give a low one if the quality is clearly lacking in their opinion. It may also suggest that users are clearer about what merits a low star rating than what merits a high star rating, though this would have to be investigated further since it cannot be assumed that all users were using the same thresholds for assigning their numeric ratings.

Finally, the neutral category stands out with its absence. Only 2 comments, in the 5 and 4.25 star category, made neutral remarks indicating that users generally prefer to say either positive or negative things if they comment at all.

4.2.2.2. Are the remarks worded in a way that is suitable to encourage volunteers rather than discourage them?

Independent of the objective quality of the work provided, one step in the analysis was to evaluate whether volunteers and organisations were using vocabulary that might encourage or discourage the original translator. In order to get insights into this, codes were given for terms that either increased or weakened a positive or negative remark. A positive remark that is increased could be considered especially positive and thereby especially motivating, while one that is weakened, might be less motivating and encouraging. On the other hand, a negative remark that is increased is stronger and therefore less discouraging while a negative remark that is weakened might be considered less discouraging. (Table 12)
Table 12 - Groups with codes and subcodes

<table>
<thead>
<tr>
<th>Group</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive, increased</td>
<td>+GOOD</td>
</tr>
<tr>
<td>Positive, weakened</td>
<td>0GOOD</td>
</tr>
<tr>
<td>Positive, double-weakened</td>
<td>00GOOD</td>
</tr>
<tr>
<td>Very positive, increased</td>
<td>+VGOOD</td>
</tr>
<tr>
<td>Very positive, weakened</td>
<td>0VGOOD</td>
</tr>
<tr>
<td>Very positive, double-weakened</td>
<td>00VGOOD</td>
</tr>
<tr>
<td>Negative, increased</td>
<td>+BAD</td>
</tr>
<tr>
<td>Negative, double-increased</td>
<td>++BAD</td>
</tr>
<tr>
<td>Negative, weakened</td>
<td>0BAD</td>
</tr>
<tr>
<td>Negative, double-weakened</td>
<td>00BAD</td>
</tr>
<tr>
<td>Negative, triple-weakened</td>
<td>000BAD</td>
</tr>
<tr>
<td>Positive, direct</td>
<td>GOOD</td>
</tr>
<tr>
<td>Very positive, direct</td>
<td>VGOOD</td>
</tr>
<tr>
<td>Neutral, direct</td>
<td>NEUT</td>
</tr>
<tr>
<td>Negative, direct</td>
<td>BAD</td>
</tr>
<tr>
<td>Very negative, direct</td>
<td>VBAD</td>
</tr>
<tr>
<td>Set aside</td>
<td>0+BAD</td>
</tr>
<tr>
<td>Set aside</td>
<td>0+GOOD</td>
</tr>
</tbody>
</table>
A remark that is both increased and weakened (for example, "overall the translation was quite good") is considered a neutral statement in regards to weakened or increased terms. For the purpose of the analysis, those categories were set aside.

The increased or weakened remarks were compared with the remarks without such increasing or weakening terms (direct terms). Additionally, the code ATT was added to the analysis separately since remarks coded as ATT should be weighed heavier than other remarks due to the fact that they constitute a remark about the person itself or their working methods rather than the output only.

4.2.2.2.1. Positive remarks

First, the direct positive remarks were compared to the increased positive, weakened positive and double-weakened positive. (Table 13, Image 13)

Within the positive remarks group, only the groups with a 5 to 4 star rating contain increased positive remarks, with the exception of 2, which could be considered not significant due to the low data volume. In the 5 and 4.75 star category, there is a significantly larger number of increased +GOOD positive remarks (0.167 and 0.368 remarks per comment) and GOOD remarks (0.217 and 0.158 remarks per comment) than weakened 0GOOD (0.067 and 0.053 remarks per comment). However, from 4.5 onwards, the number of weakened remarks is always larger than or equal to that of increased or direct positive remarks, again with the exception of the 2 star rating. In the 4 and 3.75 star rating group there are also some remarks with two weakening terms 00GOOD, which combined with the 0GOOD group form a larger number than GOOD or +GOOD remarks.

Combining the +GOOD with GOOD and 0GOOD with 00GOOD shows similar results. (Table 14, Image 14)

This suggests that only in the very high numeric rating categories 5 to 4.5 are users praising volunteers in significant numbers without weakening their remarks. Below those groups, praise is either to a similar or greater amount weakened, thereby making it potentially less motivating.

This trend becomes even clearer when the neutral GOOD is split from the increased +GOOD. (Table 15, Image 15)

Only the 5 and 4.75 star rating groups contain more remarks that are either increased or direct positive, while in all other groups there are more positive remarks that have a weakening term added than are present for the other two categories. The exception is 3.25 with an equal number of neutral GOOD and weakened 0GOOD/00GOOD remarks.

Another trend that can be observed is that only the 5 to 4 star rating groups include increased positive remarks, with the exception of the 2 star rating group which can be neglected due to the low number of overall comments in this group.
Table 13 - Positive remarks (Codes and subcodes)

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>+GOOD</td>
<td>0.167</td>
<td>0.368</td>
<td>0.200</td>
<td>0.182</td>
<td>0.143</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.250</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0GOOD</td>
<td>0.067</td>
<td>0.053</td>
<td>0.250</td>
<td>0.273</td>
<td>0.143</td>
<td>0.125</td>
<td>0.000</td>
<td>0.250</td>
<td>0.333</td>
<td>0.000</td>
<td>0.500</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>00GOOD</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.071</td>
<td>0.125</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>GOOD</td>
<td>0.217</td>
<td>0.158</td>
<td>0.200</td>
<td>0.091</td>
<td>0.143</td>
<td>0.125</td>
<td>0.000</td>
<td>0.250</td>
<td>0.167</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Image 13 - Visual representation of positive remarks (codes and subcodes)
Table 14 - +GOOD and GOOD vs. 0GOOD and 00GOOD

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>+GOOD and GOOD</td>
<td>0.383</td>
<td>0.526</td>
<td>0.400</td>
<td>0.273</td>
<td>0.286</td>
<td>0.125</td>
<td>0.000</td>
<td>0.250</td>
<td>0.167</td>
<td>0.000</td>
<td>0.000</td>
<td>0.250</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0GOOD and 00GOOD</td>
<td>0.067</td>
<td>0.053</td>
<td>0.250</td>
<td>0.273</td>
<td>0.214</td>
<td>0.250</td>
<td>0.000</td>
<td>0.250</td>
<td>0.333</td>
<td>0.000</td>
<td>0.500</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Image 14 - Visual representation of +GOOD and GOOD vs. 0GOOD and 00GOOD
### Table 15 - GOOD vs. +GOOD vs. 0GOOD and 00GOOD

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>+GOOD</td>
<td>0.167</td>
<td>0.368</td>
<td>0.200</td>
<td>0.182</td>
<td>0.143</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.250</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>GOOD</td>
<td>0.217</td>
<td>0.158</td>
<td>0.200</td>
<td>0.091</td>
<td>0.143</td>
<td>0.125</td>
<td>0.000</td>
<td>0.250</td>
<td>0.167</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0GOOD and 00GOOD</td>
<td>0.067</td>
<td>0.053</td>
<td>0.250</td>
<td>0.273</td>
<td>0.214</td>
<td>0.250</td>
<td>0.000</td>
<td>0.250</td>
<td>0.333</td>
<td>0.500</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

### Image 15 - Visual representation of GOOD vs. +GOOD vs. 0GOOD and 00GOOD
4.2.2.2. Very positive remarks

The very positive remarks category however shows a greatly different picture when increased very positive, weakened very positive, double-weakened very positive and direct very positive remarks are compared. (Table 16, Image 16)

In all categories of star rating, remarks coded as VGOOD (direct very positive) are represented in larger numbers than those with increasing or weakening terms. +VGOOD only appears in the 5 star rating group, and 00VGOOD only in the 4.75 star rating average. This suggests that while users overall, give fewer tasks a very good remark than a good one, they are more likely to commit to the very positive remark without weakening it and at the same time see less of a need to add to the very positive remark by increasing it.
Table 16 - Very positive remarks

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>+VGODD</td>
<td>0.033</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>OVGODD</td>
<td>0.017</td>
<td>0.000</td>
<td>0.000</td>
<td>0.071</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0OVGODD</td>
<td>0.000</td>
<td>0.053</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>VGOOD</td>
<td>0.600</td>
<td>0.105</td>
<td>0.100</td>
<td>0.000</td>
<td>0.143</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Image 16 - Visual representation of very positive remarks
4.2.2.2.3. Negative remarks

When increased negative, double-increased negative, weakened negative, double-weakened and triple-weakened negative as well as direct negative remarks are compared, a different trend can be observed. (Table 17, Image 17).

In the categories of negative remarks, weakened remarks appear in greater numbers than increased or double-increased remarks in nearly all star rating groups. Double-increased remarks are only present in the 4.5 group and in very small quantities (0.050 remarks per comment). Increased remarks only start to appear in significant numbers from 3.5 star rating average onwards, before that they appear in again very small numbers with 0.017 being the lowest and 0.125 the highest number of remarks per comment. When combining 0BAD, 00BAD and 000BAD, as well as +BAD and ++BAD to compare them with BAD the numbers present as follows. (Table 18, Image 18)

With the exception of the 4.25 star rating, where the combined weakened remarks are slightly lower than the BAD remarks, the users cushioned their negative remarks in all star rating groups from 5 to 3.75. BAD remarks start to be present only from 3.5 onwards, while increased BAD remarks +BAD and ++BAD are included in a higher number of remarks than the weakened or BAD ones only in the categories 2.25 and 2, both of which contain a very low number of comments (2 and 4 respectively), which is probably not enough to show a clear trend in the lower star rating categories.

It also stands out that a triple-weakened code 000BAD had to be introduced, while the positive remarks category only has a double-weakened one. This suggests that users tend to cushion negative remarks stronger than positive ones, which is also supported by the distribution of weakened and double-weakened remarks. On the other hand, there is also a ++BAD code, while there is no ++GOOD one. There are 6 star rating groups where neutral BAD remarks are present in the highest number, while for the positive remarks, there is no star rating category where the neutral GOOD is dominant.

Overall, the picture for negative remarks is not as clear as it is for positive ones. It seems that the users are split over handling negative remarks, insofar that some comments try to avoid harsh criticism and cushion more than for positive remarks while users in another group of comments seem comfortable with stating negative observations without cushioning them and are even more direct than they are for positive feedback.
### Table 17 - Negative remarks

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BAD</strong></td>
<td>0.017</td>
<td>0.000</td>
<td>0.050</td>
<td>0.091</td>
<td>0.000</td>
<td>0.125</td>
<td>0.500</td>
<td>0.500</td>
<td>0.167</td>
<td>0.250</td>
<td>0.000</td>
<td>1.000</td>
<td>0.750</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td><strong>++BAD</strong></td>
<td>0.000</td>
<td>0.000</td>
<td>0.050</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td><strong>BAD</strong></td>
<td>0.233</td>
<td>0.263</td>
<td>0.450</td>
<td>0.636</td>
<td>0.643</td>
<td>0.625</td>
<td>0.750</td>
<td>0.333</td>
<td>1.750</td>
<td>1.000</td>
<td>0.500</td>
<td>0.250</td>
<td>0.667</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td><strong>BAD</strong></td>
<td>0.017</td>
<td>0.158</td>
<td>0.650</td>
<td>0.727</td>
<td>0.143</td>
<td>0.500</td>
<td>1.000</td>
<td>0.250</td>
<td>0.500</td>
<td>1.250</td>
<td>1.500</td>
<td>0.500</td>
<td>0.500</td>
<td>2.667</td>
<td>3.000</td>
<td>0.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

### Image 17 - Visual representation of negative remarks
Table 18 - Combined 0BAD, 00BAD and 000BAD vs. combined +BAD and ++BAD vs. BAD

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0BAD, 00BAD and 000BAD</td>
<td>0.517</td>
<td>0.737</td>
<td>0.850</td>
<td>0.636</td>
<td>1.071</td>
<td>0.750</td>
<td>0.500</td>
<td>1.000</td>
<td>0.333</td>
<td>1.750</td>
<td>1.000</td>
<td>0.500</td>
<td>0.250</td>
<td>0.667</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>+BAD and ++BAD</td>
<td>0.017</td>
<td>0.000</td>
<td>0.100</td>
<td>0.091</td>
<td>0.000</td>
<td>0.125</td>
<td>0.500</td>
<td>0.500</td>
<td>0.167</td>
<td>0.250</td>
<td>0.000</td>
<td>1.000</td>
<td>0.750</td>
<td>1.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.500</td>
</tr>
<tr>
<td>BAD</td>
<td>0.017</td>
<td>0.158</td>
<td>0.650</td>
<td>0.727</td>
<td>0.143</td>
<td>0.500</td>
<td>1.000</td>
<td>0.250</td>
<td>0.500</td>
<td>1.250</td>
<td>1.500</td>
<td>0.500</td>
<td>0.500</td>
<td>2.667</td>
<td>3.000</td>
<td>0.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

Image 18 - Visual representation of combined 0BAD, 00BAD and 000BAD vs. combined +BAD and ++BAD vs. BAD
4.2.2.2.4. Very negative remarks

There were no very negative remarks that had been increased or weakened. This is partially in line with the trend observed for the VGOOD remarks, where a low number of remarks were either increased or weakened, while the majority of remarks remained as neutral VGOOD remarks. With the absence of increased or weakened remarks, the very negative category follows this trend. However, its total absence, also considering that overall there is a larger total number of very negative remarks than very positive remarks, suggests that for very negative remarks users are even more decisive than for very positive ones.

Considering the trend shown for the negative remarks, where there was a greater tendency for negative remarks to not be weakened than for positive ones, this shows that overall, users tend to be more direct with their negative remarks than with their positive ones.

4.2.2.2.5. ATT

Remarks coded as ATT were those that made negative assumptions about the volunteer or their working process, rather than commenting on the delivered task itself. Since they aim for the volunteer’s abilities at a most fundamental level, for example by comparing their work to that of a non-native speaker or of a machine translated text, remarks like this are especially likely to decrease motivation, particularly if they are untrue.

Overall, there was a surprisingly high number of such ATT remarks (14), which means that up to 8% of the comments contained such remarks. (Table 19, Image 19)

About half of the remarks were present in the categories between 2.25 and 1, and the other half in the 3.75 to 3 star rating categories. The comments in those groups remarked on the translator not being a native speaker, not having researched the subject and having hurried the task. Considering the comments, the star rating was relatively high, especially considering that the remarks in the 2.25 to 1.5 star categories marked as ATT are very similar. This raises the question whether the users commenting with ATT remarks are unnecessarily harsh with their comments, especially in the higher categories. It is however also possible that they have differing quality expectations where some would find a translation acceptable even if it was notably provided by a non-native speaker, while others would consider native speaker level the most basic requirement, for example. It is also possible that the users are not able to distinguish the mistakes a non-native speaker would make from those of a native speaker, or the errors that can be expected from machine translation from those of a human.

Independent of the link between ATT remarks and the objective quality, the relatively high number of such remarks, also in the 3.75 to 3 star rating categories, supports the observation that there is a tendency for users to be more direct with their negative comments than with their positive ones.
Table 19 - ATT remarks

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATT</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.375</td>
<td>0.000</td>
<td>0.250</td>
<td>0.500</td>
<td>0.000</td>
<td>0.500</td>
<td>0.500</td>
<td>0.000</td>
<td>0.500</td>
<td>0.667</td>
<td>1.000</td>
<td>0.000</td>
<td>0.500</td>
</tr>
</tbody>
</table>

Image 19 - Visual representation of ATT remarks
4.2.2.3. Are the remarks suited to help volunteers improve their work?

The third research sub-question regarding the comments provided by users was how well they are suited to help volunteers improve their work. For this, remarks that provided details on areas to improve in or areas that were already good, were coded as HELP. (Table 20, Image 20)

Users provided HELP remarks in all categories with the exception of 1.25, which contains only one comment and could therefore be considered not significant. The numbers of helpful remarks per comment approximately double between 2.75 and 1, compared to the 5 to 3 star rating groups.

Only a total of 5 remarks, spread across 3 comments of which 2 were in the 5 star rating average group and 1 in the 4.25 star group was related to positive remarks. Therefore, the HELP remarks were compared to the negative and very negative remarks only. (Table 21, Image 21)

The HELP remarks outweigh the very negative ones in all star rating groups, with the exception of the 1.5 star one where they are equal. On the other hand, the negative remarks appear in higher or equal numbers as the HELP remarks from the 5 to 3.5 star category and while HELP remarks then start to have higher numbers from 3.25 onwards, there are still three (2.5, 1.5 and 1) where the negative remarks are equal to the HELP ones. However, the seemingly positive balance of HELP remarks to negative remarks can be explained by the fact that the very negative remarks start to appear in greater number in the medium to lower star rating groups. This is confirmed when combining very negative and negative remarks and comparing them to the HELP remarks. (Table 22, Image 22)

In order to help volunteers with improving their work in all cases, all negative or very negative remarks should be HELP remarks, and ideally, also include positive or very positive remarks. However, while there is a large number of HELP remarks overall (157 in 171 comments), many comments contain multiple HELP remarks while others contain none. (Table 23, Image 23)

While this data includes all comments that were analysed, regardless of whether they contained a negative/very negative remark or not, it shows that users tend to leave fewer helpful remarks for volunteers that delivered tasks that were given a very high star rating (5 to 4.75), than for those with a high, medium or low star rating (4.5 and below), with the exception of 3.5 and 3 stars which might be caused by the lower number of overall comments.

A count of comments containing negative/very negative remarks without a HELP remark shows similar tendencies, albeit with more modest numbers for the 5 and 4.75 star groups. (Table 24, Image 24)

This shows that while overall, users are providing more detailed feedback when leaving negative comments than when providing positive feedback, there is still a tendency to be relatively vague especially in the higher star rating category. This would lead volunteers who are delivering tasks that are overall considered done well to be provided with less opportunities to improve further than those who are delivering poorer work. However, even in the medium range with 3.5 star rating containing 50% comments that have negative remarks without HELP, this opportunity to improve is lost for volunteers.
### Table 20 - HELP remarks

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HELP</td>
<td>0.317</td>
<td>0.526</td>
<td>1.200</td>
<td>1.364</td>
<td>1.071</td>
<td>1.000</td>
<td>0.950</td>
<td>1.750</td>
<td>1.000</td>
<td>1.250</td>
<td>2.500</td>
<td>2.000</td>
<td>1.250</td>
<td>4.000</td>
<td>3.000</td>
<td>0.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

### Image 20 - Visual representation of HELP remarks

![Image 20](image-url)
### Table 21 - HELP remarks vs. Negative and very negative remarks

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative</td>
<td>0.533</td>
<td>0.895</td>
<td>1.500</td>
<td>1.364</td>
<td>1.214</td>
<td>1.250</td>
<td>1.500</td>
<td>0.833</td>
<td>3.000</td>
<td>2.500</td>
<td>1.000</td>
<td>0.750</td>
<td>3.333</td>
<td>3.000</td>
<td>0.000</td>
<td>3.000</td>
<td></td>
</tr>
<tr>
<td>Very negative</td>
<td>0.017</td>
<td>0.000</td>
<td>0.100</td>
<td>0.091</td>
<td>0.000</td>
<td>0.250</td>
<td>0.500</td>
<td>0.750</td>
<td>0.167</td>
<td>0.750</td>
<td>0.000</td>
<td>1.500</td>
<td>1.000</td>
<td>1.333</td>
<td>3.000</td>
<td>1.000</td>
<td>2.000</td>
</tr>
<tr>
<td>HELP</td>
<td>0.317</td>
<td>0.526</td>
<td>1.200</td>
<td>1.364</td>
<td>1.071</td>
<td>1.250</td>
<td>0.750</td>
<td>1.750</td>
<td>1.000</td>
<td>3.250</td>
<td>2.500</td>
<td>2.000</td>
<td>1.250</td>
<td>4.000</td>
<td>3.000</td>
<td>0.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

### Image 21 - Visual representation of HELP remarks vs. Negative and very negative remarks
Table 22 - HELP remarks vs. Combined negative and very negative remarks

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative + Very negative</td>
<td>0.550</td>
<td>0.895</td>
<td>1.600</td>
<td>1.455</td>
<td>1.214</td>
<td>1.500</td>
<td>2.000</td>
<td>2.000</td>
<td>1.000</td>
<td>3.750</td>
<td>2.500</td>
<td>2.500</td>
<td>1.750</td>
<td>4.667</td>
<td>6.000</td>
<td>1.000</td>
<td>5.000</td>
</tr>
<tr>
<td>HELP</td>
<td>0.317</td>
<td>0.526</td>
<td>1.200</td>
<td>1.364</td>
<td>1.071</td>
<td>1.250</td>
<td>0.750</td>
<td>1.750</td>
<td>1.000</td>
<td>3.250</td>
<td>2.500</td>
<td>2.000</td>
<td>1.250</td>
<td>4.000</td>
<td>3.000</td>
<td>0.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

Image 22 - Visual representation of HELP remarks vs. Combined negative and very negative remarks
Table 23 - Number of comments containing 0, 1 or 2+ HELP remarks

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HELP count 0</td>
<td>47</td>
<td>14</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HELP count 1</td>
<td>10</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>HELP count 2+</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Image 23 - Visual representation of Number of comments containing 0, 1 or 2+ HELP remarks
### Table 24 - Comments containing negative/very negative remarks vs. HELP remarks

<table>
<thead>
<tr>
<th>Rating avg.</th>
<th>5</th>
<th>4.75</th>
<th>4.5</th>
<th>4.25</th>
<th>4</th>
<th>3.75</th>
<th>3.5</th>
<th>3.25</th>
<th>3</th>
<th>2.75</th>
<th>2.5</th>
<th>2.25</th>
<th>2</th>
<th>1.75</th>
<th>1.5</th>
<th>1.25</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Nr. of comments</td>
<td>60</td>
<td>19</td>
<td>20</td>
<td>11</td>
<td>14</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>BAD w. HELP 0</td>
<td>14</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>% of BAD w. HELP 0</td>
<td>23.33</td>
<td>36.84</td>
<td>25.00</td>
<td>9.09</td>
<td>28.57</td>
<td>12.50</td>
<td>50.00</td>
<td>0.00</td>
<td>16.67</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>25.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>BAD w. HELP 1</td>
<td>10</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BAD w. HELP 2+</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
Image 24 - Visual representation of comments containing negative/very negative remarks vs. HELP remarks
5: Discussion and recommendations

The importance of motivating all volunteers on Trommons independent of what their declared reason for volunteering is by increasing their feeling of competence and avoiding diminishing it has been shown throughout the introduction and review of the literature as well as theoretical foundations.

The analysis of the feedback data showed the trend that the majority of volunteers don’t receive written feedback on their tasks, since most comments are either left empty or they contain text other than feedback text. This means that for the majority of the volunteers, the numeric star rating alone determines to what extent the feedback increases or decreases competence motivation.

Since overall, the numeric star rating is more often given in the high to very high groups than the medium to lower ones, the feedback volunteers receive is often positive. However, the question is whether the star rating alone can be a motivating factor. For one, the relatively high number of ratings without comment with a 5-star rating in all categories suggests that while users often want to express appreciation or positive thoughts about the delivered translation, they may not put a lot of thought into the rating process but instead choose the highest in all categories as a reflex.

This is also supported by the lower but still relatively high number of ratings with an all 3-star rating, which indicates that users likely didn’t even consider rating. This relatively low interest in the feedback feature could also be observed in the high number of comments from contact persons at the organisations, where multiple tasks had received the same comment.

Even if feedback with little effort or thought was put into should be able to improve competence motivation as much as a detailed and thoughtful written comment, which would have to be further investigated, it would still not help volunteers gain competence and experience as much as a more detailed written feedback might.

Another aspect of the numeric star rating standing for itself was the observation that when comparing the comments to the star rating, there seemed to be only a rough general understanding of what kind of rating should be awarded for the different quality levels observed by the user. For example, comments that indicated that the user had made corrections to the text during the proofreading process were present not only in the 4-star rating categories, but also in the 5-star, while other users seemed to award 5 stars only if they had made no corrections at all. Since the group of users giving out star ratings isn’t fully unified in their opinions on what a 5, 4 or 3 star task looks like, the same is likely to be the case in the group of users receiving the rating. For one user, a 3-star rating might be acceptable, while for others it might indicate a much lower estimation of the quality they provided than is intended. As a result, such a rating could be lowering motivation even if this is not the intended outcome, if the star rating is not supported by a written comment which could clarify and help avoid misunderstandings.

However, also for the feedback lines included in the coding and subsequent analysis that did contain written feedback, some problematic tendencies in regards to competence and motivation could be observed. Overall, they contained a significantly higher number of negative remarks than positive ones, which was contrary to the observations made for the numeric star rating. Positive remarks were also often weakened while negative remarks were partially cushioned and partially stated directly. This trend of users being more direct in their negative statements could also be observed in the very negative
remarks and such remarks that were aimed at the user’s person or working method rather than the output. In addition, while comments often contained remarks that were suited to help volunteers improve their skills, there were still a large number of negative comments that did not give details on what the volunteer should improve. For positive remarks, details on what the volunteer should continue doing or had done well were very rare, positive remarks were often general and unspecific.

This leads to the conclusion that the vast majority of volunteers don’t receive feedback that is suited to increase their feeling of competence and thereby motivation. Only sometimes do volunteers receive feedback that is suited to help them improve their skills and rarely to help them understand what they are already doing well. Why this is the case would have to be subject of further research, but the author’s suspicion is that there are two tendencies at work: One, if a user says something positive about a task without weakening it, their professional credibility could be questioned by anyone who disagrees with their evaluation. Being critical and having high expectations are generally accepted while being overly positive could be interpreted as the user not being able to distinguish a “good” translation from a “bad” one, which makes it easier for the user to leave negative remarks than positive ones. A star rating however is less personal and therefore says less about the person who has provided it.

Another element that might be at play can be found in the theory of the fundamental attribution error, which is described in social psychology as the tendency of people to place emphasis on a person’s internal characteristics to explain their behaviour rather than external circumstances. This could explain why users are more willing to leave harsh remarks, in that it could indicate that they have painted a picture of the volunteer who delivered the task in their mind that might not reflect their actual working methods and abilities, and doesn’t take external circumstances like lack of context or reference material into consideration.

Numerous improvements could be suggested for the feedback system, for example UI changes like making the star rating groups less granular, to implementing forms where volunteers would have the opportunity to provide more detailed feedback with less effort than a written text would require. Options could also include increasing the level of information that is provided on why the feedback is important and how to give good feedback that is both, motivating and helpful.

However, the overarching observation that can be made is that users seem to have low motivation to use the feedback system as intended, which might be explained with the continuum of extrinsic motivation as described in the chapter on motivation theory. If the majority of the users have not accepted feedback and the feedback system as being valuable and important, then they would likely be closer to amotivation on the motivation continuum than to intrinsic motivation. In addition, the observation that the feedback system as it is only provides information about the task after it has been delivered, rather than ongoing support and room for improvement is relevant when considering the value of the rating system.

It is the author’s opinion that the first question that needs to be clarified is what the purpose of the feedback system is. If the intention is to evaluate and potentially filter volunteers based on their performance levels, then the current feedback system might be improved by increasing the information flow on how to use it and encouraging users to do so.
However, if the main goal is to help volunteers improve their skills and keep them motivated so they continue volunteering and maybe recruit new volunteers, the recommendation would be to consider removing the feedback system as it is entirely and replacing or at least supplementing it with a strong collaborative element. In order to increase the effort users put into feedback, they need to understand, accept and define the value of feedback. This shouldn’t happen top-down, but instead in a dialogue among the user base or between the user base and TRF. Overall, feedback should happen earlier in the process and could benefit from being more of a dialogue than a unilateral steo. In addition, starting the conversation on ways to help others improve and motivate them, could show further improvements in the future.
Reference list


