

Yes, We Care!

Results of the Ethics and Natural Language Processing Surveys

Karën Fort*, Alain Couillault†

* Université Paris-Sorbonne/EA STIH, † Université de La Rochelle/L3I
karen.fort@paris-sorbonne.fr, alain.couillault@univ-lr.fr

Abstract

We present here the context and results of two surveys (a French one and an international one) concerning Ethics and NLP, which we designed and conducted between June and September 2015. These surveys follow other actions related to raising concern for ethics in our community, including a *Journée d'études*, a workshop and the Ethics and Big Data Charter. The concern for ethics shows to be quite similar in both surveys, despite a few differences which we present and discuss. The surveys also lead to think there is a growing awareness in the field concerning ethical issues, which translates into a willingness to get involved in ethics-related actions, to debate about the topic and to see ethics be included in major conferences themes. We finally discuss the limits of the surveys and the means of action we consider for the future. The raw data from the two surveys are freely available online.

Keywords: Ethics, Moral Buffer, Whistle Blowing

Ethics in/by Natural Language Processing

Natural Language Processing (NLP), like any other science, is confronted to ethical issues, both regarding the way science is conducted (plagiarism, reproducibility, transparency) and regarding the effects of its results on society. Moreover, the 21st century is witnessing major threats to citizens' privacy in which NLP techniques play a central role. The time has therefore come to gather and share our reflections on the impact of this evolution on the way we conduct our research.

Indeed, NLP applications are often used to analyze and extract information from documents which, by their nature or their content, require thoughtful considerations regarding ethics. This is for example the case for all private content like Emails or suicide letters. It has been taken into account for a long time for medical corpora. Furthermore, the building, transformation or annotation of the (sometimes huge) language resources implies to set up and drive large scale projects involving human resources, thus generating potential issues with regards to working conditions.

On the other hand, NLP is called for when it comes to providing tools for de-identifying documents, discovering plagiarism or performing fact checking, all applications that can be considered as ethically positive.

This contrasted image of the domain has shown in the *Journée d'études* ATALA on ethics and NLP¹ and the TALN ETeRNAL workshop² that were organized last year in France.

We decided to run a survey to try to understand better the needs and concerns of the community regarding ethics.

A first poll was conducted within the French NLP community and its success encouraged us to run a second poll at the international level.

We first give a rapid overview of the range of actions performed in the more or less recent past concerning ethics and

NLP. We then detail the results of the two polls and give our interpretation of the results. Finally, we propose new short and longer-term actions to continue raising awareness and provide (at least some) answers to researchers' questions and concerns.

1. Previous Work

1.1. Ethics, Ethics Everywhere!

Identifying the main ethical issues addressed in NLP publications is an impossible task, as ethical concerns appear in lots of articles, without being mentioned explicitly in the title or keywords.

Obviously, there are exceptions, in particular articles from workshops on ethics, but also, interestingly, from people concerned by the technologies we develop. This is the case for the ethics of machine translation, which has been addressed by translators, who discuss example-based systems, underlying that they are the ones producing the precious original translations (and losing parts of their jobs), without their participation being acknowledged or rewarded (Kenny, 2011).

Some sub-domains in our field have had to face ethical issues for a long time. The most well-known case is that of speaker identification in speech processing and its usage in courts (Bonastre et al., 2003).

Ethics is also a constant preoccupation for researchers working in the biomedical domain. This is reflected in the US by the NIH Ethics Training Program³.

Moreover, it was brought to our attention that researchers who create freely available language resources and tools have not only a practical goal, but also an ethical preoccupation.

Another aspect of ethics in NLP is the very interesting research led on evaluation means and metrics, which allows to bring to light the limitations in the performance of our tools and techniques and the potential biases in their evaluation (see, for example (Gonzalo, 2010)).

¹See: http://www.schplaf.org/kf/JE_ATALA.html.

²See: <http://eternal.loria.fr/>.

³See: <https://ethics.od.nih.gov/>.

Finally, popularizing NLP applications plays an important role in improving the awareness of the public and of the authorities concerning the potential dangers and limits of NLP tools. We want here to pay a tribute to our colleague Jean Véronis, who died recently and spent a lot of time and energy making accessible to the public the subtleties of some research issues in NLP through his well-known blog⁴, which he created in 2004.

1.2. (Some) Recent Actions

Recently, some ethical concerns were raised concerning Amazon Mechanical Turk, in particular regarding the absence of a clear relation between the *Requesters* (people, including researchers, proposing a task) and the *Turkers* (workers), preventing the latter from any possibility of legal action from wrongdoings by the former (Adda and Mariani, 2010; Fort et al., 2011; Fort et al., 2014). Another issue is that of the very low wages. In order to try and improve this situation, Chris Callison-Burch proposed a tool⁵ to help *Turkers* find higher paying jobs (Callison-Burch, 2014).

The reflection started on Amazon Mechanical Turk led some of us to involve private and public bodies in the writing of an *Ethics and Big Data Charter*⁶ (Couillault et al., 2014), whose aim was to document as much as possible the building of language resources and, more generally, of data. The charter consists in a form, split into three sections, respectively dedicated to traceability, legal and licensing issues, and specific requirements (i.e. related to the very nature of the resource content). While the charter has seldom been used for what it had been designed for (i.e. document language resources), we found out that talking and publishing about it and, hence, about ethics, rose interest, if not awareness, among researchers.

As mentioned in the introduction, a *Journée d'études* ATALA and a TALN workshop were then organized on the subject of ethics and NLP, with a relatively large audience and very interesting papers.⁷

During the ETeRNAL workshop, it was suggested that, in order to try and bridge the gap between the real performance of NLP applications and the public perception of it and to share ethics-related experience and points of view, we should create a blog. The *Ethique-et-tal* blog⁸ was born. The surveys we present here are an extension of this work.

2. The Surveys

2.1. Motivations

We decided to run a survey in order to capture the point of view of NLP researchers and industrials on a number of subjects revolving around ethics. Our first goal was to

⁴See: <http://blog.veronis.fr/>.

⁵A browser extension, available here: <http://crowd-workers.com>.

⁶The charter is available here, in French and English: <http://wiki.ethique-big-data.org>.

⁷The papers (in majority in French) can be found respectively here: http://www.schplaf.org/kf/JE_ATALA.html and here: http://www.atala.org/taln_archives/ateliers/2015/ETeRNAL/.

⁸See: <http://www.ethique-et-tal.org/>.

evaluate the researchers' *moral buffer*, i.e. their attitude regarding the potential use of their research results. It then seamlessly extended to other issues.

The questions could not prevent a certain bias towards a better concern for ethics, but it did allow for disagreement to be expressed, not only in the answers themselves, but also in the free comments (for example, 17 comments were added in the French survey).

We first ran it in the French-speaking community, to check the technical and practical feasibility before extending it worldwide. For both studies, we used the LimeSurvey⁹ polling framework, a free tool¹⁰ that allows for the easy creation of anonymous on-line surveys.

The survey contained three main types of questions related to (i) the research endeavor *per se*, (ii) the effects of research results and (iii) the willingness to get involved in ethics-related actions.

Considering the high interest which arose during the French survey, we decided to drive a similar poll towards the international NLP community, using the same on-line tool, with similar questions. However, the questions differ slightly in the two polls, as we tried to improve them and their processing. We did remove questions that were not understood well by the respondents or that were too country-dependent, in particular a question regarding the notion of "fair use" (that does not exist in the French law).

Here again, the questions reflected our bias in favor of a better concern for ethics, but we are glad that people did not hesitate to express some disagreements in their answers and comments. However, it has to be noted that out of the 43 comments added to this poll, the large majority is surprisingly positive and pro-active (much more than the French ones), while a few criticize the form of the survey, i.e. the way we asked questions and the (limited) proposed answers.

The raw results of both surveys are freely available¹¹, including the answers provided by those who did not finish the surveys and which we chose not to present here.

2.2. Participation

We advertised the French survey during the JEP-TALN 2015 conference, on the LN and Corpora mailing lists and through personal Emails. Despite some connection issues, a little more than 100 persons participated in the survey between June 23rd and July 30th (we had 212 participants, but experienced connections issues, so most of them had to re-connect). To evaluate the representativeness of the sample we compared it with the number of TALN participants in recent years, which ranged from 180 to 200. We therefore consider the French survey to be representative of the French-speaking NLP (text, speech and sign) community. Out of the 102 participants who finished this first survey, 28 respondents volunteered to get involved in ethics-related actions and left their Email addresses.

We then ran a second poll aimed at the international NLP community. We advertised it on various mailing lists (LIN-

⁹See: <https://www.limesurvey.org/en/>.

¹⁰LimeSurvey is distributed under the GPL licence.

¹¹See: <http://www.schplaf.org/kf/pdf/EthicsAndNLPResults.zip>.

GUIST, Elsnest, Corpora and LN again) and, from September 8th, 263 persons participated in this survey and 200 finished it. Given that the number of ACL registered participants has been around 1,000 since 2007¹², our survey represents 20% of this community, which seems like a reasonable proportion.

Again, out of the 200 participants, 73 people declared that they are interested in participating in a working group on ethics in NLP (31 gave their Email), which is very encouraging (see Table 1).

It has to be noted that this second survey might underestimate the French-speaking community point of view, as some participants in the French poll did not participate in the international one, despite our warnings. This is why we present here a summary of the two surveys as of October 25th, 2015.

3. Results

The international survey returned results which did not differ much from the ones from the French survey, which lets us think that the question of ethics is quite independent of local laws, rules or habits.

One question, though, returned different results from one survey to the other. To the question "Have you ever refused a project due to ethical issues?", 40.20% of the respondents in the French survey selected "Yes" while they were 23.50% in the international survey (see Figures 1 and 2).

As mentioned by some of the French survey respondents in their comments to this question, major efforts have been made at the European level to take ethical issues into account when submitting projects for proposal. For example, an ethics self-assessment form¹³ has to be attached to all projects submitted for funding. This might be one reason for the differences between the French and international results.

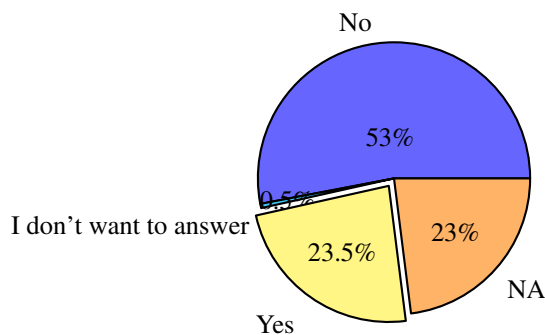


Figure 1: "Have you ever refused a project due to ethical issues?" (international survey).

3.1. The Researcher, the Politician and the Citizen

A large majority of researchers, in both surveys, consider that both citizens and politicians are not aware of the lim-

¹²Personal communication with P. Rasmussen, ACL Business Manager, on October 21st, 2015.

¹³See: http://ec.europa.eu/.../1660136-1645175-h2020_-_guidance_ethics_self_assess_en.pdf.

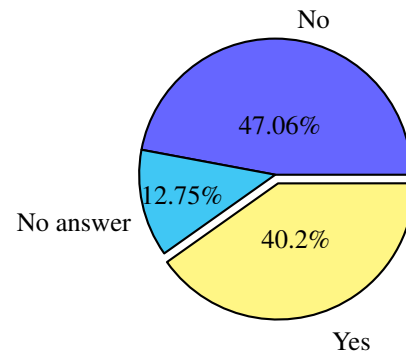


Figure 2: "Avez-vous déjà refusé ou limité un projet pour des raisons éthiques ? / Have you ever refused a project due to ethical issues?" (French survey).

its and possibilities of the tools we create. To the question "Do you think the public is aware of the limits and possibilities of the tools we create?", a large majority answer "No" (see Figures 3 and 4). To the same question concerning the authorities (government, politicians, etc), a little less large majority answer "No". This is quite worrying as the former are directly concerned by applications which can potentially violate their privacy, sometimes on the authorities' demand.

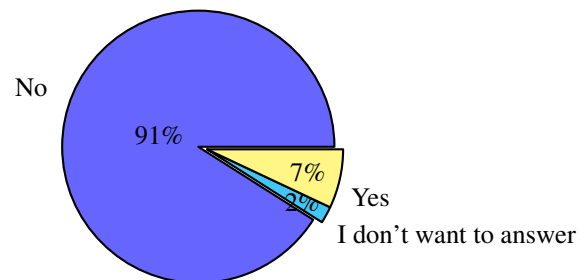


Figure 3: "Do you think the public is aware of the limits and possibilities of the tools we create?" (international survey).

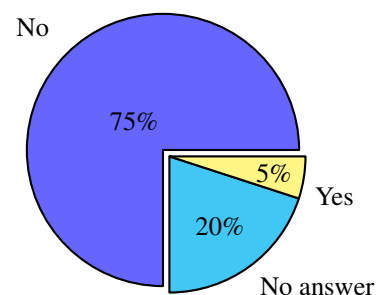


Figure 4: "Pensez-vous que le grand public est conscient des limites des capacités des outils de TAL ? / Do you think the public is aware of the limits and possibilities of the tools we create?" (French survey).

3.2. Ethics in the Researcher's Everyday Life

3.2.1. Exploiting Research Results

A short majority of participants in the surveys consider that they are responsible for the usages of their applications (see

Survey	French	International
Date of the survey	July 2015	September 2015
Number of participants	(212 -) 102	(263 -) 200
Number of participants to TALN / ACL	200	1,000
Number of declared volunteers	28	73

Table 1: Participation in the two surveys.

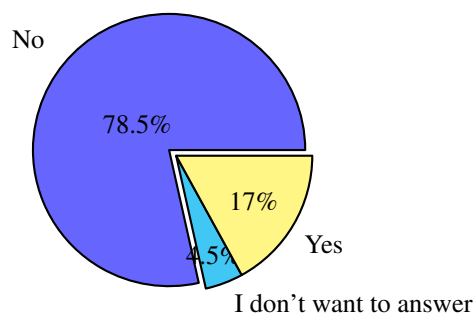


Figure 5: "Do you think the authorities are aware of the limits and possibilities of the tools we create?" (international survey).

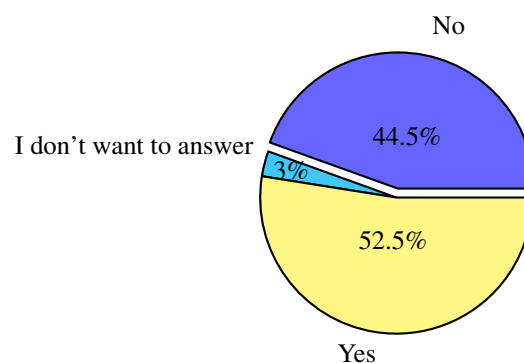


Figure 7: "Do you consider yourself responsible for the usages imagined from the applications/algorithms you create?" (international survey).

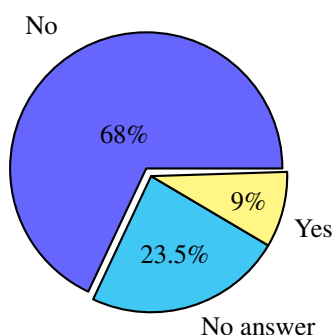


Figure 6: "Pensez-vous que les pouvoirs publics sont conscients des limites des capacités des outils de TAL ? / Do you think the authorities are aware of the limits and possibilities of the tools we create?" (French survey).

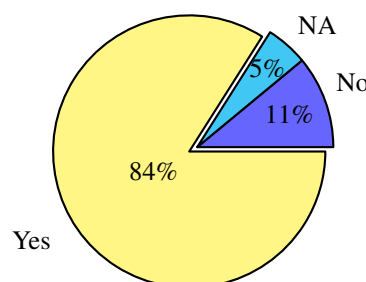


Figure 8: "In your projects, do you consider the licensing and distribution of your language data?" (international survey).

Figure 7 and Table 2).

This means that, although there is a moral buffer (and therefore room for awareness raising), researchers do care about the potential misuse of their results. However, some comments, especially in the French survey, show that a minority of researchers is doubtful about our role and responsibility, doubts which can be summarized with this quote "Let's bother about scientific issues and leave ethics in the hands of the users."¹⁴. Obviously, this key question would require a wider debate focused on NLP results exploitation.

3.2.2. Documenting Results

Regarding the researchers "day-to-day" behavior in their research project, the study confirms the results we witnessed in (Couillault et al., 2014) with a real concern for licensing issues (see Figures 8 and 9).

However, although a majority of the participants declare

¹⁴"Occupons-nous des questions scientifiques et laissons les questions éthiques aux utilisateurs."

that they know how the data producers were remunerated in the projects they were involved in (see Figures 10 and 11), we showed in (Couillault et al., 2014) that this information is almost never published. One explanation for this might be that they do not consider this as relevant for a scientific publication, so that encouraging researchers to do so could help lift the veil on our behavior as employers.

3.2.3. Training

Few academic programs seem to include a training on ethics (less than 20% in both surveys, with around 20 that are detailed in the comments), especially in France (less than 15%). This shows that there is room for improvement and sharing of good practices and that France is lagging behind on the subject.

3.2.4. Blowing the Whistle

A clear majority (63.00%) of the international participants (the question was not included in the French survey) declare that they do not know how to blow the whistle in

Oui, c'est tout à fait mon rôle / Yes, it is definitely my role	40.20%
C'est un rôle partagé par l'ensemble de l'équipe / It is a role which is shared by the whole team	34.31%
C'est le rôle d'un des membres de l'équipe / It is the role of one of the team members	0.98%
Mon sujet de recherche n'est pas pertinent pour cette question / My research subject is not relevant to this question	10.78%
Non, ce n'est pas la responsabilité du chercheur / No, it is not the researcher's responsibility	23.53%

Table 2: "Vous considérez-vous responsable des utilisations faites des outils que vous développez ? / Do you consider yourself responsible for the usages imagined from the tools you develop?" (French survey).

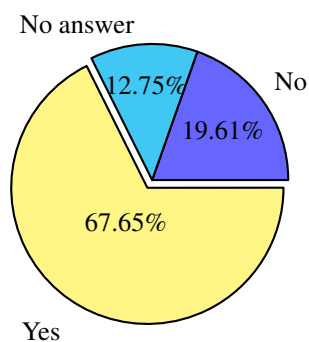


Figure 9: "Dans vos projets intégrez-vous dès le départ la possibilité de pérenniser et redistribuer vos données ? / In your projects, do you consider the licensing and distribution of your language data?" (French survey).

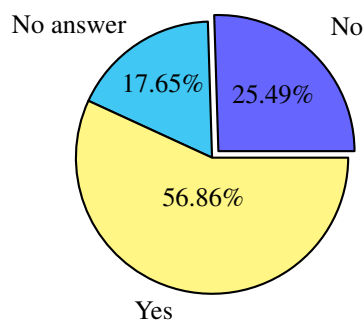


Figure 11: "Dans les projets auxquels vous avez participé, savez-vous comment les producteurs de données ont été rémunérés ? / In the projects in which you participated, do you know how the data producers were remunerated?" (French survey).

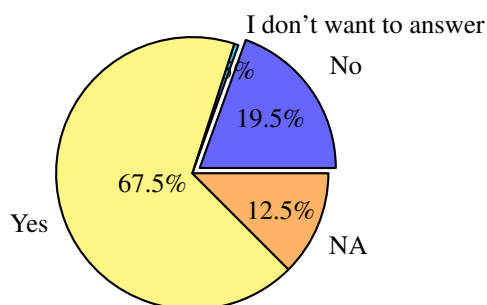


Figure 10: "In the projects in which you participated, do you know how the data producers were remunerated?" (international survey).

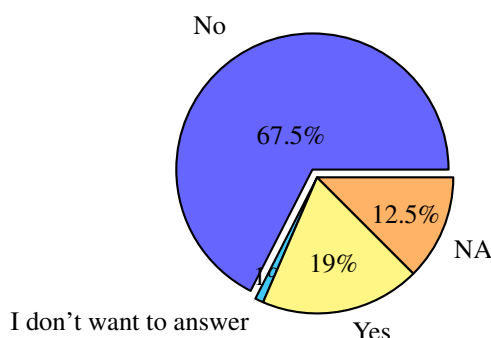


Figure 12: "Is there any class on ethics in the curricula in which you participate?" (international survey).

case they witness a major ethical infringement in the domain (see Figure 14).

3.3. "I do not want to answer"

Respondents in the international survey had the possibility to choose "I do not want to answer". The ratio of respondents who chose this option varies greatly from one question to the other, ranging from 7.5% for the question "Are you ready to participate in a working group on ethics in NLP?" to 0% to the question "In your projects, do you consider the licensing and distribution of your language data?". It seems that the more a question requires self-commitment, or to give an opinion, the higher the ratio of "I do not want to answer". This might be due to the fact that respondents are reluctant to talk about their opinion (and maybe to share it on-line).

4. Going Further

4.1. Discussion

We realize now that some of the questions in both surveys lacked context and were interpreted in different ways than expected (and probably differently by respondents). In particular, it seems that our question on whistle blowing ("Do you know of any way to blow the whistle (issue an alert) in case of a major ethical infringement in your domain?") was understood differently than expected. We have to admit that we had in mind government-related procedures¹⁵, whereas some comments from the respondents who answered "Yes" show that they understood it as ethical

¹⁵For example in the US: <http://www.whistleblowers.gov/>, or in the UK <https://www.gov.uk/whistleblowing/what-is-a-whistleblower>.

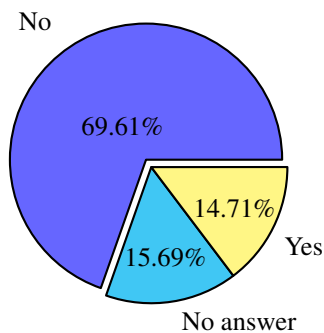


Figure 13: "Existe-t-il une sensibilisation à l'éthique dans les formations dans lesquelles vous intervenez ? / Is there any class on ethics in the curricula in which you participate?" (French survey).

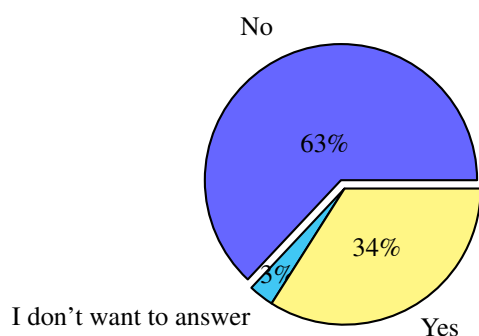


Figure 14: "Do you know of any way to blow the whistle (issue an alert) in case of a major ethical infringement in your domain?" (international survey).

issues in colleagues' behavior (which is interesting too, but very different from our interpretation).

A question in the French survey, concerning the need for a "fair use" à la française was not understood by some of the respondents. We therefore do not present it here.

Some comments in the international survey criticize the black-and-white possible answers we imposed in the international poll ("Yes", "No" and "I do not want to answer" for most questions) in order to simplify the processing of the survey.

Finally, the international participation rate seems fair, but more participants would obviously have been more satisfactory.

Overall, these issues are all linked to our hastiness in creating the surveys and running them in order to be able to advertise the French one during JEP-TALN 2015 and to have the international results ready for LREC. We believe that this time pressure is one of the (if not the) major causes of ethical issues in research, and that we should find ways to resist to this attraction, individually, and as a community.

4.2. What's next?

There seems to be a high demand for action, as shown by the answers to the question "Do you think that ethics should be part of the subjects in the call for papers" (see Figures 15 and 16) and the relatively high number of people ready to participate to a working group (30 and 73 persons in the

French and international polls, respectively). This encouraged us to move forward and take further actions in cooperation with other colleagues.

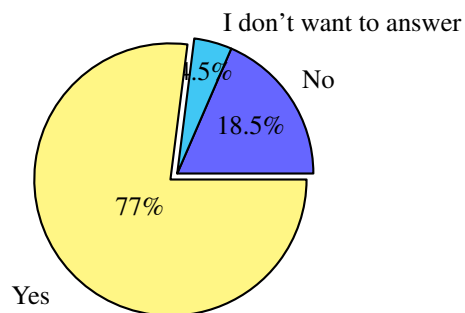


Figure 15: "Do you think that ethics should be part of the subjects in the calls for papers of the conferences in our domain?" (international survey).

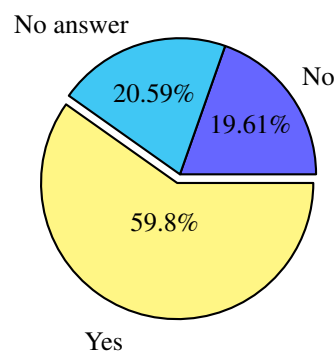


Figure 16: "Pensez-vous que l'éthique doit faire partie des sujets de l'appel général de la conférence TALN ? / Do you think that ethics should be part of the subjects in the calls for papers of the conferences in our domain?" (French survey).

The Ethics and NLP blog we created is a first means of action, which allows researchers and industrials to express themselves on the subject of ethics and to popularize our real achievements and their limits. In parallel, we created a mailing list for French-speaking people interested in the subject, which will be enlarged to a wider community as soon as possible.

We have also worked on a second version of the *Ethics and Big Data Charter* and extended it to data sets pertaining to other domains or with specific constraints such as medical, health data or personal data.

Finally, our proposal of a special issue of the TAL journal on Ethics and NLP has been accepted and a workshop dedicated to ethics, ETHI-CA² (ETHics In Corpus collection, Annotation and Application) will be held during LREC 2016.

Ethics should be shaped by the community itself, so let us gather and talk about it. Someone added the following comment in the international survey "I think the community is mature enough to have this conversation today". We do too! The LREC conference seems to us (one of) the perfect place(s) to start this discussion.

Acknowledgements

We want to thank here our colleagues who participated in the many reflections on ethical issues during the *Journée d'études*, the ETeRNAL workshop and after, in particular Gilles Adda (LIMSI-CNRS/IMMI), Maxime Amblard (University of Lorraine/LORIA), Jean-Yves Antoine (University of Tours/LI) and Hugues de Mazancourt (Eptica). We also thank the LREC reviewers, who gave us some useful and precise advice to improve this paper. Part of the research presented here was made possible thanks to fundings from the French Ministry of Culture, through the DGLFLF (délégation générale à la langue française et aux langues de France).

Appendix

Table 3 summarizes the results for each survey. Note that the second question (concerning whistleblowing) was not asked in the French survey. Moreover, the possible answers to the first question were different in French:

- Oui, c'est tout à fait mon rôle / Yes, it is definitely my role (40.20%)
- C'est un rôle partagé par l'ensemble de l'équipe / It is a role which is shared by the whole team (34.31%)
- C'est le rôle d'un des membres de l'équipe / It is the role of one of the team members (0.98%)
- Mon sujet de recherche n'est pas pertinent pour cette question / My research subject is not relevant to this question (10.78%)
- Non, ce n'est pas la responsabilité du chercheur / No, it is not the researcher's responsibility (23.53%)

We regrouped the first three answers into "Yes" in the following table and we did not put the fourth one.

Also note that "No answer" corresponds to slightly different answers in the two polls. In the French survey, this covers the cases where LimeSurvey did not register any answer, except in the first question, where it corresponds to "My research subject is not relevant to this question". In the international survey, this corresponds to "I don't want to answer" and "N/A" answers (when applicable).

5. Bibliographical References

Adda, G. and Mariani, J. (2010). Language resources and amazon mechanical turk: legal, ethical and other issues. In *Proc. of LREC Legal Issues for Sharing Language Resources workshop*, Valletta, Malta, May.

Bonastre, J.-F., Bimbot, F., Boë, L.-J., Campbell, J. P., Reynolds, D. A., and Magrin-chagnolleau, I. (2003). Person authentication by voice : A need for caution. In *Proc. of Eurospeech 03*.

Callison-Burch, C. (2014). Crowd-workers: Aggregating information across turkers to help them find higher paying work. In *Proc. of the AAAI Conference on Human Computation and Crowdsourcing (HCOMP-2014)*, November.

Couillault, A., Fort, K., Adda, G., and De Mazancourt, H. (2014). Evaluating Corpora Documentation with regards to the Ethics and Big Data Charter. In *Proc. of LREC*, Reykjavik, Iceland, May.

Fort, K., Adda, G., and Cohen, K. B. (2011). Amazon Mechanical Turk: Gold mine or coal mine? *Computational Linguistics (editorial)*, 37(2):413–420, June.

Fort, K., Adda, G., Sagot, B., Mariani, J., and Couillault, A. (2014). Crowdsourcing for Language Resource Development: Criticisms About Amazon Mechanical Turk Overpowering Use. In Zygmunt Vetulani et al., editors, *Human Language Technology Challenges for Computer Science and Linguistics*, pages 303–314. Springer International Publishing, July.

Gonzalo, J. (2010). Benchmarking and evaluation campaigns: the good, the bad and the metrics. In *LREC 2010 Workshop on Methods for the automatic acquisition of Language Resources and their evaluation methods*, Valletta, Malta.

Kenny, D. (2011). The ethics of machine translation. In *New Zealand Society of Translators and Interpreters Annual Conference 2011*, Auckland, New Zealand, 4-5 June.

Question	Yes/Oui (%)	No/Non (%)	No answer (%)
Do you consider yourself responsible for the usages imagined from the applications/algorithms you create? Vous considérez-vous responsable des utilisations faites des outils que vous développez ? (see note above)	52.50 75,49	44.50 23.53	3.00 10.78
Do you know of any way to blow the whistle in case of a major ethical infringement in your domain?	34.00	63.00	3.00
Are you ready to participate in a working group on ethics in NLP? Etes-vous d'accord pour participer à un groupe de travail sur l'éthique dans le TAL ?	32.00 26.47	60.50 38.24	7.50 35.29
Have you ever refused a project due to ethical issues? Avez-vous déjà refusé ou limité un projet pour des raisons éthiques ?	23.50 40.20	53.00 47.06	23.50 12.75
In your project, do you consider the licensing and distribution of your language data? Dans vos projets intégrez-vous dès le départ la possibilité de pérenniser et redistribuer vos données ?	84.00 67.65	11.00 19.61	5.00 12.75
In the projects in which you participated, do you know how the data producers were remunerated? Dans les projets auxquels vous avez participé, savez-vous comment les producteurs de données ont été rémunérés ?	67.50 56.86	19.50 25.49	13.00 31,18
Do you think the public is aware of the limits and possibilities of the tools we create? Pensez-vous que le grand public est conscient des limites des capacités des outils de TAL ?	7.00 4.90	91.00 75.49	2.00 19.61
Do you think the authorities are aware of the limits and possibilities of the tools we create? Pensez-vous que les pouvoirs publics sont conscients des limites des capacités des outils de TAL ?	17.00 8.82	78.50 67.65	4.50 23.53
Is there a class on ethics in the curricula in which you participate? Existe-t-il une sensibilisation à l'éthique dans les formations dans lesquelles vous intervenez ?	19.00 14.71	67.50 69.61	13.50 15.69
Do you think ethics should be part of the list of subjects in the calls for papers of the conferences in our domain? Pensez-vous que l'éthique doit faire partie des sujets de l'appel général de la conférence TALN ?	77.00 59.80	18.50 19.61	4.50 20.59

Table 3: Summary of the answers given in both surveys by the respondents who finished the survey, in the order of the international survey (see warnings above).