Crowdsourcing for language learning
Some considerations from deontological or consequentialist ethics

Jean-Yves Antoine
LI, Université François Rabelais, Tours

enetCollect, september 7th, 2017
Some ethical thoughts ... from a NLP researcher

Not an ethicist
Simply a NLP researcher with ethical concerns

Natural Language Processing
• Applications to information retrieval and assistive technologies for disabled people

Initial questioning: augmentative and alternative communication

- Word prediction for virtual keyboard
- Severe motion impairments (tetraplegia...)

Increase of input speed
Learning of linguistic competencies
Ethical researches: aims

- The better intents can have their negative counterparts...
- Systematic assessing the **ethical impact** of new digital technology
Crowdsourcing for Language Learning material

- Ethical involvement of the crowd
- Research methodology
- Deontological Ethics

Language Learning Solutions

- Ethical impact of the applications
- Research outcomes
- Consequentialist Ethics
Ethics and action

**Law**  
*What is allowed and what is forbidden*
- External (state) regulation
- Must be strictly respected

**Ethics**  
*What is good/right and what is bad/wrong*
- Duty: self-regulation based on moral judgments
- Elaborated by individuals or some community
- Guide for personal/institutional conduct

Different approaches of ethics

- **Deontological Ethics**
- **Virtue Ethics** (ancient Greece)
- **Consequentialist Ethics**
- **Applied Ethics** (analysis of particular issues)
Ethics: approaches

Deontological Ethics

- **Principle first** – An action must respect some moral principles / rules
- **Ethics** – Debates on the definition of these rules

**Example**

Consequentialist Ethics

- **Consequence first** – Teleological approach: right actions have good outcomes

**Examples**
- Utilitarianism (Jeremy Bentham, John Stuart Mills)
- Hans Jonas (1990) Imperative of responsibility → Precautionary principle
Crowdsourcing for Language Learning material

Ethical principles have been proposed on labour activities, IT or knowledge production and access

Deontological Ethics

Language Learning Solutions

Usages of new technologies are hardly predictable: how to follow some moral rules

[Moore, 1999]

Consequentialist Ethics
Crowdsourcing

Deontological Ethics → PAPA Principles

PAPA Principles

P Privacy
A Accuracy
P Property
A Accessibility

Crowdsourcing for Language Learning material

Paid crowdsourcing

[Durward, Blohm, Leimeister 2016]
Crowdsourcing: Privacy

**Ability of the individual to personally control information about oneself**

**Privacy and crowdsourcing** – Crucial ethical issues on (paid) crowdsourcing platforms, which collect usually a large amount of personal data

- Which personal data?
- Which use and control on these data?
- Respect of the regulations on the anonymization / control of data
- European Data Protection Directive 95/46/CE (under revision)
- Privacy goes beyond biographical or medical data and personal opinions/preferences
- Individual quality metrics can be seen as an invasion into the privacy of workers [Kajino and al. 2014]

**Question** – Will performance data be associated to every crowdworker?
Crowdsourcing: Accuracy

**Extent to which data are correct, reliable and certified** [Wang & Strong 1996]

**Accuracy and crowdsourcing** – Central concern of crowdsourcing, but accuracy extends the standard definition of reliability as considered by data science

**enetCollect**

**“Pure” Accuracy (Data Accuracy)**
- Training
- Task definition
- User interface
- Reliability metrics

Influence of individual characteristics
- Location (region)
- Age, book reading frequency
- Openness, conscientiousness

**Privacy**

**Beyond data accuracy**

**Transparency**

Must we really look for experts?
Crowdsourcing: Accuracy

**Extent to which data are correct, reliable and certified** [Wang & Strong 1996]

**Accuracy and crowdsourcing** – Central concern of crowdsourcing, but accuracy extends the standard definition of reliability as considered by data science.

**enetCollect**

**Pure accuracy**
- Training
- Task definition
- User interface
- Reliability metrics
- Workers profiling

**Transparency** [Agerfalk & Fitzgerald 2008]
- **Honesty**: explicit vs. implicit (direct vs. indirect) – Notify workers that a free labour is performed and for which aim
  - **Counter example**: reCAPTCHA [Lung 2016]
- **Model transparency** – Workers must be able to understand how their contribution is used in the final data
  - **weighted or majority score? consensus dispute?** [Doan et al. 2011]

**Question** – Explicit vs. implicit crowdsourcing? Which combination model?
Crowdsourcing: Property

*Intellectual property* — *Patrimonial and moral property*

**Accuracy and crowdsourcing** — Collective property, shared between the crowdworkers and the scientists that designed the research and combined the data

**enetCollect**

---

**Patrimonial**
- Right to use?
- Right to sell?
- Right to generate incomes?

**Moral**
- Citation of every crowdworker?
- Right to modify

---

**Question** — Which economical model? Which copyright / distribution licence?
Crowdsourcing: Accessibility

**Extent to which data are attainable by the mass of people**
Technical, economical but also cognitive perspective (data illiteracy, understanding)

**Accuracy and crowdsourcing** – Data accessibility but also accessibility to the crowdsourcing procedure: maximising the number and the diversity of participants

**enetCollect**

**Data accessibility**
- User-friendly interface, tutorials
- Distribution licence

**Property**

**Privacy + Accuracy**

**Crowdsourcing accessibility**
- Equality = maximising the crowd
  - Recruitment strategy
  - “no a priori identification” (K. Fort)
- Help the user to contribute
  - Community management, gamification
  - User-friendly interface [Doan et al. 2011]
  - Recommendation tools [Schnitzer et al. 2015]
  - Task adaptation to users’ skills
Crowdsourcing: conclusion

Some decisions that will impact the ethical value of the project

- Privacy
- Accuracy
- Property
- Accessibility

Explicit / implicit crowdsourcing
Crowdworkers behaviour evaluation
Crowdworkers profiling / open call
Standard crowdsourcing / deep citizen science
Model for merging contributions
Distribution licence / commercial use
Language Learning Solutions

Consequentialist Ethics

Consequences of the developed solutions

- (Industrial) Risk
- Systematic check-up

Risk ontology
- NLP, HMI & assistive technologies
  [Lefeuvre-Halftermeyer et al 2016]
Risk


Risk – Effect of uncertainty on objectives

Effet – Deviation from the expected — positive and/or negative

Objective – Objectives can have different aspects (such as financial, health and safety, and environmental goals) → ethical issues

Risk

positive effect

Systems Evaluation

negative effect

Consequentialist Ethics
Risk analysis

<table>
<thead>
<tr>
<th>Risk Source</th>
<th>Vulnerability</th>
<th>Criticity</th>
</tr>
</thead>
</table>
| Element which alone or in combination has the intrinsic potential to give rise to risk | Something having a susceptibility to a risk source ⇒ Element that should be impacted | $C = LR \times L$
| LR Level of risk (magnitude) | L Likelihood (probability) |

Is the application a risk source?  
Who / what endures the potential risk?  

Typology of risk sources

enetCollect
Typology of risk sources

Risk sources organised according to vulnerability

Risk is multifactorial – Any risk source should concern several objects of vulnerability, while a combination of sources should give rise to a unique risk

⇒ Ontology with graphs rather than a pure hierarchical taxonomy
Risks on individuals

- **Individual**
  - Physical risk
  - Cognitive risk
  - Psychological risk

- **2. Society**
  - 2.1. Risk on human interaction
  - 2.2. Risk on social system
Physical risk

Modification of the physical integrity of an individual or any element

- **Physical risk**
- **impact on individual**
- **impact on physical environment**
- **Physical impact**
- **Physiological impact**

Logistics: voice control for order pickers
- Productivity increase
  - lombago
  - musculo-skeletal disorders

[Govaere and Wioland 2016]
Physical risk : enetCollect

Modification of the physical integrity of an individual or any element

- Physical risk
- Impact on individual
  - Impact on physical environment
- Physical impact
  - Physiological impact

enetCollect
Modification of a cognitive function or of the general cognitive state

Cognitive functions – mental functions of the ICD-10 (International Classification of Diseases) of the World Health Organisation (WHO/OMS)
Cognitive risk

Modification of a cognitive function or of the general cognitive state

Information retrieval and cognitive memory: **Google effect** [Sparrow et al. 2011]

The repetitive use of a search engine leads to sensible modifications of long-term memory

- we are less likely to remember details we believe will be accessible online
- people's ability to learn information offline remains the same

Computer-aided writing — word prediction, automatic translation, orthographic correction...

- Decrease of the cognitive stimulation: impact on language abilities?
- Long-term tool dependence: lack of autonomy?
Cognitive risk

Modification of a cognitive function or of the general cognitive state

Cognitive risk → Cognitive ability → Cognitive delay → Language

enetCollect: language abilities

Language learning applications

• **Evaluation** – Does the application allow an increase of the linguistic abilities
• **Autonomy** – Do we observe a cognitive tool dependence?
• **Learning bias** – Does the application favors one ability against others (example: communication skills vs. grammatical mastery)
Psychological risk

Temporary or permanent modification of the psychological state

Psychological risk

Permanent effect

Mood disorder

Anxiety disorder

Personality disorder

Temporary effect

Psychological fatigue

Stress / emotionnal shock

Speech synthesis for disabled people

⇒ tracheotomy, degenerative disease

⇒ New voice, or preventive recording
Psychological risk

Temporary or permanent modification of the psychological state

Psychological risk
  Permanent effect
    Mood disorder
    Anxiety disorder
    Personality disorder
  Temporary effect
    Psychological fatigue
    Stress / emotionnal shock

(Self) training
  New technologies can be a source of stress
  Training difficulties: feeling of inferiority
Social risk

- Individual
  - Physical risk
  - Cognitive risk
  - Psychological risk

- Society
  - Risk on social link
  - Risk on the social system
Risk on social link

Impact on the inter-individuals relations as well as the social insertion

Social link
- Relation with technology
- Inter-individual relations
  - Individual rights and freedoms
    - Privacy & E-reputation
    - Legal responsibility
    - Labour law
    - Freedom of expression
    - Intellectual property

Authorship authentication

Crowdsourcing
Illicit work [Sagot et al. 2011]
Risk on social link

Impact on the inter-individuals relations as well as the social insertion

- Social link
  - Relation with technology
  - Inter-individual relations
  - Individual rights and freedoms
    - Privacy & E-reputation
    - Legal responsibility
    - Labour law
    - Freedom of expression
    - Intellectual property
Risk on the social system

Modification of the social system: politics, economy, culture...

Social System
- Political risk
- Economical risk
- Sociocultural risk

Automatic translation: “Google-ese”
- English as a pivot language
- “natural languages could progressively evolve to seamlessly integrate the linguistic biases of algorithms” [Kaplan 2014]

Search engine
- Paid links, ads: search queries = auction
- (Key)words get an economic value
  - Linguistic capitalism: English [Kaplan 2014]
Risk on the social system

Modification of the social system: politics, economy, culture...

Social System

- Political risk
- Economical risk
- Sociocultural risk

- Which language register (colloquial or literary) will be used by crowdworkers?
- Which learning aims (communication-first or grammar-first) will be promoted by the language learning solutions?

Example: disappearance of encyclopaedia publishers with the emergence of Wikipedia

What will be the economical impact of eLearning solutions on humane teaching?
Some concerns on individuals as well as society
Jean-Yves Antoine
LI, Université François Rabelais de Tours, Blois

Jean-Yves.Antoine@univ-tours.fr

Multidisciplinary work

Anaïs Lefevre-Halftermeyer (NLP)
LIFO, U. Orléans

Willy Allègre (Assistive Technologies & occupational therapy)
CMRFF de Kerpape, Ploemeur

Virginie Govaere (ergonomics, occupational risk)
LIFO, INSA Bourges
References: deontological approach


J. Lung (2016) Ethical and legal considerations of Recaptcha.

R.O. Mason (1986) Four ethical issues of the information age, Mis Quarterly, pp. 5-12.


References: consequentialist approach


