

Università degli Studi di Torino

**Multilinguisme et variétés
linguistiques en Europe à l'aune de
l'intelligence artificielle**

**Multilinguismo e variazioni
linguistiche in Europa nell'era
dell'intelligenza artificiale**

**Multilingualism and Language
Varieties in Europe in the Age of
Artificial Intelligence**

Édité par, a cura di, edited by

Rachele Raus, Università di Bologna
Alida Maria Silletti, Università di Bari
Silvia Domenica Zollo, Università di Verona
John Humbley, Université de Paris



**UNIVERSITÀ
DI TORINO**

Special Issue - 2022

De Europa

***Multilinguisme et variétés linguistiques en Europe
à l'aune de l'intelligence artificielle***

***Multilinguismo e variazioni linguistiche in Europa
nell'era dell'intelligenza artificiale***

***Multilingualism and Language Varieties in Europe
in the Age of Artificial Intelligence***

Édité par, a cura di, edited by

Rachele Raus, *Università di Bologna*

Alida Maria Silletti, *Università di Bari*

Silvia Domenica Zollo, *Università di Verona*

John Humbley, *Université de Paris*



**UNIVERSITÀ
DI TORINO**

Special Issue - 2022

De Europa

European and Global Studies Journal

www.deeuropa.unito.it

Collane@unito.it

Università di Torino

ISBN ebook: 9788875902179

ISBN cartaceo: 9788855268431



Quest'opera è distribuita con
Licenza Creative Commons Attribuzione.
Condividi allo stesso modo 4.0 Internazionale.
Copyright © 2022, stampa 2023



Ledizioni 
The Innovative LEDpublishing Company

Ledizioni LediPublishing
Via Antonio Boselli, 10
20136 Milano – Italia
www.ledizioni.it
info@ledizioni.it

With the support of the
Erasmus+ Programme
of the European Union



Jean Monnet Chair
*The EU in a Challenging
World*



In cooperation with:



Introduction/ Introduzione / Introduction

Rachele Raus

7

Première partie : réflexions et études de cas

Introduction

Réflexions et études de cas à l'aune de l'intelligence artificielle.
Vers de nouveaux observables linguistiques ?

John Humbley, Silvia Domenica Zollo

35

Quelques réflexions sur le multilinguisme à l'aune de l'intelligence artificielle

Intelligence artificielle et langues minoritaires : du bon ménage ?
Quelques pistes de réflexion

Giovanni Agresti

47

Elaborazione automatica dei linguaggi diversi dall'inglese:
introduzione, stato dell'arte e prospettive

Guido Vetere

69

Études de cas

Enabling additional official languages in the EU for 2025
with language-centred Artificial Intelligence

Kepa Sarasola, Itziar Aldabe, Nora Aranberri

91

Langages et savoirs : intelligence artificielle et traduction
automatique dans la communication scientifique

Maria Luisa Villa, Maria Teresa Zanola, Klara Dankova

107

Variation et traduction

Terminologie, intelligence artificielle, psychologie cognitive :
réflexions sur les interactions possibles dans l'étude de la variation
en langue spécialisée

Anne Condamines

131

| | |
|---|-----|
| Human-machine interaction: how to integrate plain language rules in the revision cycles of Neural Machine Translation output <i>Christopher Gledhill, Maria Zimina</i> | 149 |
| A Journey in Neural Machine Translation <i>Philippe Langlais</i> | 173 |
| Deuxième partie : expérimentations pédagogiques | |
| <hr/> | |
| <i>Introduction</i> | |
| Expérimentations pédagogiques : perception et utilisation de l'intelligence artificielle dans la formation universitaire <i>Alida Maria Silletti, Rachele Raus</i> | 199 |
| Le multilinguisme européen et l'IA. Enquête auprès des futurs décideurs <i>Dardo de Vecchi</i> | 215 |
| Les dispositifs de traduction automatique et la recherche terminologique comme outils pédagogiques pour des étudiant·e·s en droit <i>Francesca Bisiani</i> | 247 |
| Fraseologia, traduzione e <i>digital literacy</i> nel contesto universitario: riflessioni e proposte per un percorso didattico sperimentale <i>Silvia Domenica Zollo, Silvia Calvi</i> | 263 |
| Variation terminologique et traduction automatique : une expérience didactique dans l'enseignement du français sur objectif spécifique (FOS) <i>Jana Altmanova, Luca Bottiglieri</i> | 285 |
| Les genres textuels | |
| La traduzione automatica neurale: uno strumento di sensibilizzazione per la formazione universitaria in lingua e traduzione francese <i>Ilaria Cennamo, Maria Margherita Mattioda</i> | 307 |

| | |
|--|-----|
| Artificial Intelligence and Machine Translation: perceptions, opinions and experiences of Italian Graduate Students of English as a Foreign Language <i>Alessandra Molino</i> | 337 |
| Assessing the efficacy of machine translation across genres <i>Chiara Abbadessa, Monica Albini, Elisa De Paoli, Francesca Del Nobile</i> | 355 |
| Intelligenza artificiale e traduzione automatica nel contesto della formazione universitaria di lingua tedesca <i>Lucia Cinato</i> | 365 |
| Annexes | 385 |

Introduction

Rachele Raus

Ce numéro spécial de la revue *De Europa* a été élaboré à partir d'une réflexion interdisciplinaire et multilingue qui a été menée dans le cadre d'une recherche sur les droits, le multilinguisme et les variétés linguistiques en Europe à l'aune de l'IA¹ que je coordonne à l'intérieur du projet *Artificial Intelligence for European Integration*², dirigé par Umberto Morelli, promu par le Centre d'études européennes *To-EU* de l'Université de Turin³ et co-financé par la Commission de l'Union européenne.

Notre propos était de réfléchir plus généralement sur les conséquences négatives et/ou positives de l'IA sur les variétés linguistiques et le multilinguisme, ce dernier étant une valeur de l'UE. La recherche a débuté le 6 octobre 2020 lors de l'intervention de Laurent Romary⁴, directeur du Comité 37 de l'Organisation internationale de normalisation (ISO), à l'intérieur d'un premier colloque international organisé en distanciel à l'Université de Turin par le groupe pilote du projet. Ensuite, le 23 avril 2021, nous avons eu l'occasion d'organiser un deuxième colloque en distanciel portant sur la thématique spécifique de notre intérêt. Les interventions présentées à ce colloque sont désormais disponibles dans le site du projet⁵. La journée a été suivie par une matinée de séminaire (*workshop*) où les participant.e.s venant de plusieurs universités italiennes et françaises ont eu l'occasion de présenter leurs travaux inédits sur la didactique des langues étrangères et l'IA.

Rachele Raus, Università di Bologna, rachele.raus@unibo.it

¹ Cette recherche a regroupé jusqu'à aujourd'hui environ une centaine de personnes venant de plusieurs pays de l'UE, du Canada et du Brésil.

² Voir le lien <http://www.jmcoe.unito.it/home> (dernière date de consultation de tous les sites cités : le 20 octobre 2021).

³ <https://www.dcps.unito.it/do/home.pl/View?doc=toeu.html>

⁴ <https://media.unito.it/?content=9806>

⁵ <http://www.jmcoe.unito.it/>

Compte tenu de l'intérêt du débat soulevé, nous avons décidé de recueillir quelques-unes des contributions des deux journées du 23 et 24 avril 2021, dûment retravaillées à la suite des nombreuses rencontres des groupes de travail et des personnes concernées, et d'y ajouter des essais de collègues qui ont voulu participer à cette initiative, en enrichissant un débat déjà passionnant. Par conséquent, ce que nous proposons par ce numéro est un livre inédit qui peut se vanter d'être parmi les premiers à s'occuper de ce type de thématique, du moins en Europe.

1. La question initiale

Dès les années 1990, plusieurs organisations internationales, notamment l'Unesco⁶, ont commencé à se poser la question de la « fracture numérique » causée, entre autres, par les technologies actuelles qui sont plus diffusées dans certains pays que dans d'autres. Sans compter d'autres risques du même genre : par exemple, le fait que les langues orales ont moins la possibilité de devenir visibles dans la toile ; que les grands ensembles de corpus et des données (*big data*) sont de plus en plus gérés par des entreprises privées⁷ ; que certains pays, comme le Canada ou, en Europe, la France et l'Allemagne, ont investi massivement dans la production de pages publiées dans la toile et également dans l'industrie des langues ; que la plupart des données linguistiques et des corpus sont normalement disponibles en langue anglaise ; que souvent les outils informatiques multilingues utilisent l'anglais comme langue pivot (voir la contribution de Vetere dans ce numéro), etc.

Les groupes de travail⁸ que nous avons organisés autour de ces questionnements ont ouvert plusieurs pistes de recherche possibles :

1. la question des droits linguistiques par rapport à la possibilité de certains pays d'investir dans l'industrie des langues aux dépens des autres. Cette thématique, soulevée par des linguistes et des juristes des groupes de travail, a été abordée par le Rapporteur spécial des Nations unies sur les questions relatives aux minorités, Fernand de Varennes⁹,

⁶ Cf. <http://www.unesco.org/new/index.php?id=50219>

⁷ A propos de la gestion des données par rapport à l'IA, cf. l'intervention de l'informaticien Ciro Cattuto au colloque du 6 octobre 2020 au lien : <https://media.unito.it/?content=9806>

⁸ Les groupes ont été divisés par langue véhiculaire (FR, IT, EN).

⁹ https://drive.google.com/file/d/1YPSFX-v9ITB4N1tj_eFOI7Sinc9NBgcY/view

- lors du colloque du 24 avril 2021. Le débat se poursuivra en 2022, lors d'un colloque organisé par l'Université Catholique de Lille et prévu pour le printemps de l'an prochain;
2. la nécessité de sélectionner attentivement les sources et les données utiles pour entraîner les algorithmes d'IA.

Cette deuxième question en soulève deux autres qui lui sont étroitement liées : d'abord, la relation que ces données entretiennent avec les formes de discrimination, ce que Bartoletti a appelé la « discrimination algorithmique » (Bartoletti 2020) ; ensuite, la relation entre ces données et des sources « d'autorité », ce dont on a commencé à discuter avec Laurent Romary lors du colloque du 6 octobre 2020. En effet, le critère de source d'autorité proposé par l'ISO (2009 : 7) recommande que cette source soit officielle ou garantisse des critères de qualité, ce qui implique, de manière indirecte, la présence d'une source institutionnelle. C'est sans doute l'une des raisons pour laquelle les tous premiers entraînements des algorithmes neuronaux utilisés dans les logiciels de traduction automatique ou dans les dispositifs numériques ont été faits sur les corpus multilingues des organisations internationales. Une autre en est que celles-ci disposent d'une grande quantité de données multilingues déjà disponibles et prétraitées pour le traitement automatique. C'est ce qui est arrivé pour le traducteur de Google¹⁰, dont les réseaux neuronaux d'apprentissage profond se sont entraînés sur les corpus des Nations unies (ONU 2017 : 120), ainsi que pour Facebook, qui a utilisé les documents multilingues du Parlement européen pour entraîner ses traducteurs automatiques (Le Cun 2019 : 277), ou encore pour le concordancier Linguee¹¹, dont les corpus fournissent le matériel d'entraînement au traducteur automatique DeepL¹², corpus qui viennent des institutions de l'UE, de l'ONU, du gouvernement canadien... Cela dit, l'utilisation de corpus et de données plus variés, mais non institutionnels, qui donc ne sont pas forcément vérifiés ni forcément de qualité, a fini par améliorer les performances des outils informatiques par rapport à d'autres qui utilisent des corpus institutionnels.

Pendant le colloque d'octobre, nous avons justement cité l'exemple de l'accord des noms et des adjectifs au féminin. En effet, ce problème, qui

¹⁰ <https://translate.google.com>

¹¹ <https://www.linguee.fr>

¹² <https://www.deepl.com/fr/translator>

caractérise avant tout les noms de métiers et de profession, s'aggrave lors de la rédaction ou de la traduction automatique des textes faite par des outils dont les réseaux de neurones d'apprentissage profond s'entraînent sur des corpus institutionnels qui, privilégiant des textes juridiques et/ou politiques, finissent par utiliser le « neutre » masculin pour renvoyer aux catégories d'acteurs. Cela produit des erreurs d'accord du féminin lors de la traduction automatique des langues romanes, comme il arrive pour le traducteur de Google ou DeepL, qui sont parmi les outils les plus utilisés, entre autres, par les universitaires italiens (voir les contributions dans la deuxième partie de ce numéro).

Ce problème pourrait être résolu, du moins en partie¹³, en utilisant des corpus variés qui ne sont pas forcément vérifiés, et donc non institutionnels, qui seraient plus proches des usages linguistiques et de la langue courante.

Le questionnement sur les données et sur leurs critères de sélection a conduit à s'interroger sur la possibilité de prévoir des critères en amont de l'entraînement des réseaux neuronaux. L'intervention humaine par rapport à la « machine » se déplacerait donc du travail actuel (en aval) de révision des textes produits par celle-ci au travail préalable (en amont) de réflexion et d'implémentation des critères linguistiques et discursifs utiles pour bien entraîner l'IA (sélection de sous-corpus choisis exprès pour la tâche spécifique ; mise en place de contraintes qui permettraient d'améliorer et de faciliter l'apprentissage supervisé de la machine, etc.). Ces critères permettraient de réaliser concrètement les recommandations du livre blanc européen de 2020 par rapport au contrôle humain lors de la conception de l'outil afin d'éliminer des discriminations linguistiques et de genre et d'« imposer des contraintes opérationnelles » favorisant le multilinguisme et les variations linguistiques (Commission européenne 2020 : 22 et 25).

Il s'agit sans aucun doute d'une solution onéreuse, qui demande une réflexion préalable sur ce qui fait « autorité » par rapport à l'entraînement de l'IA et qui implique également tout un travail d'étiquetage de corpus de

¹³ Bien que le manque de féminisation de certains noms de métiers et de profession caractérise plus généralement plusieurs langues romanes, entre autres le français et l'italien, cette tendance empire dans les corpus institutionnels qui entraînent les algorithmes et qui recourent normalement au masculin « neutre » pour renvoyer aux acteurs dans les discours juridiques et ensuite administratifs (voir aussi Marzi 2021).

grande taille pour les rendre plus utiles par rapport à l'apprentissage. Cela dit, il s'agit d'un travail lourd mais nécessaire si l'on veut que l'IA devienne réellement un outil au service du multilinguisme. En effet, ce n'est pas parce qu'un traducteur automatique est multilingue qu'il peut favoriser le multilinguisme si, par exemple, il utilise l'anglais comme langue pivot, comme il arrive pour la plupart des outils de l'industrie des langues actuelle.

Outre la question des droits et des données, les groupes de travail sur le multilinguisme et l'IA se sont aussi intéressés à un volet pédagogique de la recherche, ce qui a permis de promouvoir des travaux en classe centrés sur les langues et l'IA à travers l'utilisation d'outils de traduction automatique exploitant des réseaux neuronaux d'apprentissage profond (voir la deuxième partie de ce numéro).

2. D'une réflexion critique à une IA qui favorise le multilinguisme

Bien que plusieurs contributions de ce numéro finissent par être critiques par rapport aux conséquences de l'IA sur le multilinguisme (voir, entre autres, les articles d'Agresti et de Vecchi dans ce numéro) et par rapport à un marché, tel celui de l'industrie des langues, qui aujourd'hui donne une fausse perception des langues comme de simples moyens de communication, voire des « codes » interchangeables, il faut admettre que c'est justement en questionnant l'IA que nous avons pensé à des possibilités concrètes de l'utiliser de manière différente et donc en faveur du multilinguisme.

Cela nous a permis de revenir sur la question des données de manière nouvelle, en proposant de concevoir et de fabriquer une application multilingue qui faciliterait la communication inclusive et qui serait destinée aux administrations publiques. Cette application serait d'abord conçue pour la langue italienne et ensuite pour d'autres langues européennes (en français, puis en allemand et enfin en espagnol et/ou en anglais, en tenant compte de la variation diatopique de ces langues). Elle sera réalisée avec le soutien de l'École polytechnique de Turin à partir de critères qui puissent développer une IA en faveur des variétés linguistiques et du multilinguisme. Nous espérons présenter l'application italienne lors d'une journée d'études que nous souhaitons organiser pour la fin de 2023 et qui permettrait de présenter également des initiatives similaires à la nôtre,

comme le nouveau traducteur automatique que Facebook a inauguré en 2020 et qui n'utilise plus l'anglais comme langue pivot¹⁴, ou comme le transcritteur IA Azur de Microsoft qui est utilisé par le Parlement européen et qui tient compte des différents accents régionaux ou nationaux (Aeles 2020). Cette journée montrerait la possibilité d'améliorer les performances de l'IA grâce au choix de critères discursifs et linguistiques qui tiennent compte des variations diaphasiques, diatopiques et diastratiques des langues.

Par rapport au volet pédagogique, les groupes de travail ont mis en place, pendant l'année académique 2020-2021, une expérimentation dont les résultats sont présentés dans la deuxième partie de ce numéro. Cette expérimentation a permis de mettre en relief comment une formation ciblée permet de développer une conscience critique par rapport à l'IA, ce qui favorise une meilleure utilisation des outils de l'industrie des langues qui s'appuient sur l'IA.

Cette réflexion ne porte pas atteinte à la confiance initiale que les étudiant.e.s déclarent faire à l'IA, mais tend à la renforcer grâce à une meilleure compréhension de ces outils et de leur potentiel.

Au final, ce numéro de la revue *De Europa* entend favoriser l'amélioration de l'IA d'une part, en privilégiant des critères linguistiques et discursifs précis fixés en amont de son apprentissage et d'autre part, en permettant une compréhension majeure de la manière dont elle privilégie certaines données de sortie. De cette manière, les articles ci-rassemblés entendent également contribuer aux produits de recherche finaux du projet, qui sont attendus pour 2023, à savoir :

- des recommandations à l'intention des décideurs européens pour l'élaboration de politiques qui favorisent la mise en œuvre d'une intelligence artificielle respectueuse du multilinguisme ;
- un vade-mecum, destiné au personnel travaillant dans des domaines professionnels (technique, informatique, organismes de normalisation, ...), qui contiendra des suggestions pour concevoir des algorithmes d'IA qui favorisent le multilinguisme et les variations diatopiques et diastratiques des langues, en se focalisant notamment sur les langues officielles de l'UE.

¹⁴Cf. <https://www.20minutes.fr/high-tech/2889599-20201020-facebook-met-point-outil-traduction-automatique-entre-100-langues-differentes>

3. Présentation des contributions

Ce numéro, qui contient des contributions trilingues (IT, FR, EN) résumant les points forts de la première année de recherche des groupes de travail, se pose comme le premier d'une trilogie, chaque livre renvoyant aux résultats d'une année de travail. Nous avons voulu privilégier une approche de vulgarisation des contenus scientifiques de manière à joindre un public très large, sans rien enlever à la rigueur scientifique de nos propos.

De même, nous avons essayé de favoriser une écriture inclusive, sans pour autant obliger les personnes qui ont participé à cette publication à privilégier des choix linguistiques spécifiques. C'est la raison pour laquelle, dans certaines contributions, on a privilégié la féminisation (par le point médian, par les tirets ou par des parenthèses) et que dans d'autres (la majorité des cas) on a choisi d'utiliser le masculin soi-disant « neutre ».

Par rapport à la structure du numéro, nous l'avons divisé en deux parties : dans la première, nous avons voulu présenter des réflexions générales sur l'IA par rapport au multilinguisme et aux variétés linguistiques. Ensuite, nous avons regroupé les contributions qui portent sur des études de cas spécifiques, surtout des cas concernant la traduction automatique et l'IA, qui constituent la toute dernière sous-section de la première partie.

Dans la seconde section du numéro nous avons rassemblé les articles concernant les résultats des expérimentations pédagogiques menées par des universités et des écoles professionnelles françaises et italiennes qui ont participé au projet : la Kedge Business School, l'EDEH, l'EM-Normandie, la Montpellier Business School et l'École supérieure de commerce de Clermont, l'Université Catholique de Lille, l'Université Catholique du Sacré-Cœur de Milan, l'Université de Naples — L'Orientale, l'Université de Turin et l'Université de Vérone. Nous avons également pu intégrer à cette section des articles ou des propos venant des étudiant.e.s universitaires (niveau Master 1-2 ou 3^e cycle), étant donné que le projet prévoyait une méthode de recherche-action.

Avant de laisser nos lectrices et nos lecteurs à la lecture du numéro, nous tenons à remercier John Humbley, Alida Maria Silletti et Silvia Domenica Zollo pour leur travail infatigable et ponctuel de relecture et de mise en forme des articles, ainsi que pour avoir nourri la réflexion commune qui a permis à ce travail de voir le jour.

Nous remercions également Umberto Morelli, qui coordonne le projet *Artificial Intelligence for European Integration* et qui a toujours su gérer au mieux tous les groupes de travail, même lorsque les conditions pour mener à bien le projet sont devenues difficiles à cause de la pandémie à COVID-19.

Enfin, merci à toutes les personnes qui ont contribué de manière plus ou moins directe à nourrir le débat sur ces thématiques passionnantes.

Nous espérons qu'il pourra se poursuivre et devenir de plus en plus riche, l'IA étant désormais devenue un élément incontournable de notre vie quotidienne.

Bibliographie

- Aeles Joris (2020). *Rapprocher le Parlement européen des citoyens grâce à l'IA*.
<https://pulse.microsoft.com/fr-be/transform-fr-be/government-fr-be/fa1-rapprocher-le-parlement-europeen-des-citoyens-grace-a-lia/>
- Bartoletti Ivana (2020). *An Artificial Revolution. On Power, Politics and AI*. Edimbourg : Indigo.
- Commission européenne (2020). *Livre blanc. Intelligence artificielle. Une approche européenne axée sur l'excellence et la confiance*. COM(2020) 65 final.
- Le Cun Yann (2019). *Quand la machine apprend*. Paris : Odile Jacob.
- Marzi Eleonora (2021). « La traduction automatique neuronale et le biais de genre : le cas des noms de métiers entre l'italien et le français ». *Synergies Italie*, 17, 19-36. <https://gerflint.fr/Base/Italie17/marzi.pdf>
- Organisation des Nations unies (2017). *Change. Rapport annuel 2017*. Genève : Nations unies.
- Organisation internationale de normalisation — ISO (2009). *Norme 23185. Assessment and benchmarking of terminological resources*. Genève : ISO.

Introduzione

Rachele Raus

Il presente numero speciale di *De Europa* è il frutto di una riflessione interdisciplinare e multilingue maturata attorno a diversi eventi organizzati nell'ambito del panel concernente i diritti e le variazioni linguistiche in Europa nell'era dell'intelligenza artificiale¹ da me diretto all'interno del progetto *Artificial Intelligence for European Integration*², promosso dall'Università di Torino (direzione del prof. Umberto Morelli) e cofinanziato dalla Commissione europea.

L'interrogativo iniziale che abbiamo voluto sollevare è se l'IA potesse avere un impatto negativo sulle varietà linguistiche e sul multilinguismo, valore "aggiunto" dell'UE³, o se potesse, e in che modo, divenire utile per la promozione di essi. La ricerca ha dato modo di organizzare un primo convegno a distanza il 6 ottobre del 2020 al quale è stato invitato il prof. Laurent Romary⁴ quale rappresentante del Comitato tecnico 37 dell'ISO e successivamente un secondo convegno, svoltosi a distanza il 23 aprile 2021, i cui interventi sono disponibili nel sito del progetto⁵ e al quale è seguita una giornata seminariale (*workshop*) in cui sono stati presentati dei lavori inediti sulla didattica e l'IA svolti in numerose università italiane e francesi.

Dato l'interesse del dibattito sollevato, si è scelto di raccogliere alcuni dei contributi proposti in queste giornate, opportunamente rielaborati

Rachele Raus, Università di Bologna, rachele.raus@unibo.it

¹ Il panel ad oggi ha coinvolto un centinaio di persone circa provenienti da vari paesi dell'UE, dal Canada e dal Brasile. Questo senza considerare il contributo fornito dalla componente studentesca che è stata coinvolta nelle ricerche sulla didattica.

² Maggiori dettagli al link <http://www.jmcoe.unito.it/home> (tutti i link della presentazione sono stati consultati il 20 ottobre 2021).

³ Così è infatti spesso definito proprio nei siti istituzionali europei, in particolare del Parlamento.

⁴ <https://media.unito.it/?content=9806>

⁵ <http://www.jmcoe.unito.it/>

anche a seguito dei numerosi incontri avuti con molte delle persone che vi hanno partecipato, integrandoli con articoli di colleghe e colleghi che hanno voluto arricchire ulteriormente un dibattito già fruttuoso. Conseguentemente, ne è risultato un volume interamente inedito, che può dirsi tra i primi ad affrontare, almeno in Europa, questo tipo di tematiche.

1. L'interrogativo iniziale

Dagli anni 1990, diverse organizzazioni internazionali, e in particolare l'Unesco⁶, hanno cominciato ad affrontare il problema del divario di conoscenza tra i vari paesi a seguito del fatto che le tecnologie attuali, a cominciare da Internet, siano più diffuse in alcuni paesi piuttosto che in altri. La presenza di lingue orali che hanno meno la possibilità di divenire visibili nel web, la gestione dei grandi dati, anche linguistici, che sono sempre più in mano a colossi privati⁷, la capacità e la volontà di alcuni paesi come il Canada, o, in Europa, la Francia o la Germania di investire massivamente nella produzione di pagine web e nei prodotti dell'industria linguistica rispetto ad altri che hanno meno disponibilità economiche, la disponibilità maggiore dei dati in lingua inglese, la concettualizzazione della traduzione multilingue a partire dall'inglese veicolare (cfr. il contributo di Vetere nel presente volume), ecc. sono tutte questioni che certamente si ripercuotono sul multilinguismo e sulla questione delle lingue minoritarie che sarà trattata in questo volume.

Dalle prime riflessioni dei gruppi di lavoro⁸ che abbiamo costituito attorno a queste tematiche sono scaturiti cinque assi di ricerca privilegiati che abbiamo voluto ulteriormente indagare. Il primo concerne la capacità dei paesi di investire nell'industria linguistica e più generalmente le politiche linguistiche implementate dai diversi paesi dell'UE per rispettare i diritti linguistici. Questo primo asse, che ha coinvolto persone specialiste nell'ambito linguistico e giuridico, è stato introdotto dall'intervento del prof. Fernand de Varennes⁹, relatore speciale delle Nazioni Unite, nell'ambito dei lavori del 24 aprile 2021 e proseguirà con i lavori promossi dall'U-

⁶ Cfr., ad esempio, <http://www.unesco.org/new/index.php?id=50219>

⁷ A proposito della gestione dei *big data* in relazione all'IA, cfr. l'intervento di Ciro Cattuto al convegno del 6 ottobre 2020, disponibile al link <https://media.unito.it/?content=9806>

⁸ I gruppi di lavoro sono stati suddivisi sulla base delle lingue veicolari utilizzate (FR, IT, EN).

⁹ https://drive.google.com/file/d/1YPSFX-v9ITB4N1tj_eFOI7Sinc9NBgcY/view

niversité Catholique di Lille che sta attualmente organizzando un convegno su tali tematiche per la primavera del 2022.

Il secondo asse di ricerca mira a riflettere sulla necessità di selezionare con attenzione le fonti e i dati utili per l'addestramento dell'IA, e si correla direttamente ad altri due degli assi proposti, quello sulla relazione tra tali dati e la "discriminazione algoritmica" (Bartoletti 2020) e quello sulla relazione tra i dati e la nozione di fonte "autorevole", sulla quale si è cominciato a discutere con il prof. Romary durante il convegno del 6 ottobre 2020. In effetti, il criterio della fonte autorevole dell'ISO (ISO 2009: 7) raccomanda che tale fonte sia ufficiale o garantisca criteri di qualità, cosa che implica indirettamente la presenza di una fonte di tipo istituzionale. Questo è probabilmente uno dei motivi per cui, assieme al fatto di disporre di grandi dati utili, i primi addestramenti degli algoritmi neurali di apprendimento profondo utilizzati per i *tool* di traduzione automatica o nell'industria linguistica si sono basati sui corpora delle organizzazioni internazionali. Così è avvenuto per il traduttore di Google¹⁰, rispetto alle fonti delle Nazioni Unite (ONU 2017: 120), per gli algoritmi di traduzione automatica utilizzati da Facebook rispetto alla documentazione del Parlamento europeo (Le Cun 2019: 277), per l'addestramento di software di concordanze come Linguee¹¹, i cui corpora forniscono il materiale per il traduttore automatico DeepL¹², rispetto ai testi del governo canadese e/o delle istituzioni dell'UE e dell'ONU... Tuttavia, l'utilizzo di corpora e dati di tipo misto, non per forza istituzionale né per forza verificati e perciò di qualità, ha finito per dare risultati migliori rispetto alle performance dell'IA utilizzata in questi strumenti. Proprio durante il convegno di ottobre, citavamo, ad esempio, il fatto che la variante di genere (es. la declinazione al femminile), già problematico più generalmente nell'italiano e nelle lingue romanze soprattutto in relazione ai titoli e nomi di professione, viene ulteriormente a perdersi in testi frutto di elaborazione (redazione o traduzione) di strumenti supportati dall'IA addestrata su corpora istituzionali, essenzialmente a carattere politologico-giuridico, che utilizzano massivamente il maschile "neutro" o "non marcato" per indicare delle categorie di attori. Non c'è da sorprendersi, perciò, che testi declinati al femminile in francese diventino esclusivamente maschili nella traduzione ita-

¹⁰ <https://translate.google.it/?hl=it>

¹¹ <https://www.linguee.it/>

¹² <https://www.deepl.com/it/translator>

liana come ancora avviene per traduttori automatici diffusi come Google o DeepL, peraltro tra i più usati nel mondo universitario italiano come si può evincere dalla seconda parte del presente volume. Questo potrebbe essere evitato, almeno in parte¹³, in caso di utilizzo di corpora misti non per forza verificati, e non istituzionali, ma più indicativi rispetto agli usi linguistici della lingua comune.

Questi interrogativi sui dati e sui loro criteri di selezione hanno portato a promuovere la possibilità di inserire correttivi e criteri a monte dell'apprendimento delle reti neurali, non limitando perciò l'intervento umano alle attività di *postediting* dei testi prodotti in modo automatico, come avviene per lo più oggi, ma prevedendolo a monte rispetto all'apprendimento della macchina. In tal senso, i correttivi consentirebbero di realizzare concretamente i criteri indicati nel libro bianco europeo del 2020 di una sorveglianza durante la fase di progettazione dell'IA per eliminare a monte possibilità di discriminazione linguistica e di genere (Commissione europea 2020: 21) e d'implementare "regole operative" (*ibidem*: 24) atte a promuovere il multilinguismo e la variazione linguistica.

Certamente, si tratta di una via più lunga da percorrere, che porta a riflessioni sui corpora "autorevoli" da selezionare e soprattutto sui criteri da adottare in fase di apprendimento supervisionato, nonché implica la necessità di annotare corpora ampi perché diventino abbastanza significativi da poter essere utili per l'addestramento dell'IA, ma si tratta anche di un lavoro necessario se si vuole davvero che essa diventi un vettore a sostegno del multilinguismo. Precisiamo, infatti, che un traduttore automatico multilingue non è di per sé, solo per il fatto di essere multilingue, un supporto al multilinguismo se processa i dati favorendo, ad esempio, l'inglese veicolare che resta maggioritariamente lingua ponte nell'industria linguistica attuale.

Infine, il quinto asse di ricerca si è focalizzato più specificatamente sulla didattica e ha permesso di promuovere in classe lavori centrati sulle lingue e l'IA tramite l'ausilio di strumenti di traduzione automatica basati su algoritmi di apprendimento profondo e quindi con intelligenza artificiale.

¹³ Sebbene la mancanza di femminilizzazione di molti titoli e nomi di professione caratterizzi più generalmente le lingue romanze, tra le quali il francese e l'italiano, tale tendenza peggiora nei corpora istituzionali che addestrano gli algoritmi e che ricorrono solitamente al maschile "non marcato" o "neutro" per rinviare agli attori nel discorso giuridico e amministrativo (cfr. anche Marzi 2021).

2. Da una riflessione critica all'IA come vettore a sostegno del multilinguismo

Malgrado alcuni degli interventi presenti in questo volume siano animati da uno spirito di analisi critica rispetto agli attuali utilizzi degli algoritmi nell'industria linguistica e a un mercato delle lingue che spinge inesorabilmente verso una concettualizzazione di esse come di meri strumenti di comunicazione (cfr. gli articoli di Agresti e di de Vecchi all'interno del numero) e quindi veri e propri codici privi di spessore culturale e simbolico, va detto che proprio da tale analisi è sorta la possibilità concreta di un utilizzo diverso e virtuoso dell'IA a favore di un multilinguismo reale.

In tal senso, ad esempio, le ricerche effettuate dai nostri gruppi sugli assi concernenti i dati e il loro utilizzo per l'addestramento di reti neurali stanno confluendo verso la concettualizzazione e la realizzazione di un applicativo utile alla comunicazione inclusiva e rivolto alle pubbliche amministrazioni. Tale applicativo, che sarà dapprima realizzato per la lingua italiana, poi anche per altre lingue europee (anzitutto il francese, a seguire il tedesco e probabilmente anche l'inglese e/o lo spagnolo), sarà sviluppato con il contributo del Politecnico di Torino¹⁴, a partire da criteri linguistici e discorsivi che possano favorire lo sviluppo di un'IA a sostegno della variazione linguistica e del multilinguismo. Il nostro desiderio è che questa iniziativa possa esser presentata in una giornata di studi assieme ad altre consimili, come ad esempio quella del nuovo traduttore automatico inaugurato da Facebook¹⁵ nel 2020 e che non utilizza più l'inglese ponte o come il trascrittore IA Azur di Microsoft che è usato dal Parlamento europeo e che tiene conto degli accenti regionali o nazionali (Aeles 2020). Tali iniziative mostrerebbero come, partendo da criteri discorsivi e attenti alla variazione delle lingue, *in primis* quelle diatopica e diastratica, sia possibile migliorare le performance dell'IA e favorirne conseguentemente un utilizzo realmente virtuoso a sostegno del multilinguismo e della variazione linguistica.

Parimenti, rispetto all'asse cinque, l'esperimento didattico intrapreso nel corso dall'A.A. 2020-2021 ha permesso di mettere in luce come tramite la didattica sia possibile far maturare un senso di riflessione critica nei confronti dell'IA che porti a un utilizzo migliore degli strumenti dell'industria linguistica supportati da reti neurali. Tale riflessione non mina quel

¹⁴ Il team di sviluppo è coordinato dalla prof.ssa Tania Cerquitelli, che ringraziamo.

¹⁵ Cfr. <https://www.20minutes.fr/high-tech/2889599-20201020-facebook-met-point-outil-traduction-automatique-entre-100-langues-differentes>

senso di fiducia iniziale nei confronti dell'IA che la componente studentesca spesso nutre, come vedremo nella seconda parte del presente volume, ma va a rafforzarlo a fronte di una comprensione maggiore di essa e delle sue potenzialità.

Il presente numero di *De Europa*, quindi, vuole contribuire alla riflessione sull'IA al fine di favorire le sue capacità a supporto del multilinguismo, e ciò grazie da un lato all'addestramento a partire da criteri linguistici e discorsivi precisi stabiliti a monte, dall'altro, consentendo a un pubblico giovane e/o non esperto di acquisire una maggior comprensione e consapevolezza del perché essa finisca per privilegiare determinati output. In tal senso, il volume vuol essere un tassello verso l'elaborazione dei due prodotti finali del progetto che sono attesi per il 2023, ovvero una guida a destinazione dei decisori e delle deciderici dell'UE rispetto alle politiche multilingui da adottare in relazione all'IA e il vademecum a destinazione di un pubblico professionale (enti di normazione, attori dell'industria linguistica...) per sviluppare strumenti e quadri normativi capaci di favorire un'IA a reale supporto della variazione linguistica e del multilinguismo.

3. La strutturazione del numero

Il presente numero speciale presenta contributi trilingui (IT, FR, EN) relativi al primo anno di ricerca dei gruppi di lavoro. Si tratta del primo volume di una trilogia che nei prossimi anni riassumerà i risultati principali delle ricerche svolte nel triennio.

Quanto agli articoli qui raccolti, si è volutamente privilegiato un approccio di tipo divulgativo in modo che essi, seppur supportati da argomentazione e prove scientifiche, potessero rivolgersi a un pubblico vasto. Per lo stesso motivo, si è cercato di favorire una scrittura che fosse inclusiva, senza però obbligare coloro che hanno contribuito al numero a operare scelte precise e univoche in tal senso. Per questo, in alcuni contributi si è scelto di femminilizzare usando il punto mediano, le parentesi o i trattini, in altri (la maggior parte degli articoli) si è lasciato il maschile "non marcato".

Abbiamo deciso di strutturare il testo in due sezioni: nella prima, abbiamo introdotto una riflessione generale sull'IA rispetto alla variazione linguistica e al multilinguismo (cfr. i contributi di Agresti e di Vetere) e abbiamo illustrato dei casi di studio specifici (cfr. i contributi di Sarasola

et alii e di *Villa et alii*). Entrando sempre più all'interno di casi di studio, l'analisi della traduzione automatica è stata quella che ha maggiormente destato l'attenzione degli addetti e delle addette ai lavori (cfr. gli articoli di Condamines, Gledhill e Zimina, Langlais).

Nella seconda parte del volume, abbiamo raccolto gli articoli concernenti la sperimentazione didattica condotta dalle università e scuole professionali che hanno partecipato alla nostra iniziativa (per questo volume: la KEDGE Business School, l'EDEH, l'EM-Normandie, la Montpellier Business School et l'École supérieure de commerce de Clermont, l'Université Catholique de Lille, l'Università Cattolica del Sacro Cuore di Milano, le Università L'Orientale di Napoli, le Università di Torino e di Verona). In tale sezione, compaiono anche direttamente o indirettamente i contributi di studentesse e studenti che hanno partecipato al progetto (livello magistrale e dottorato), dato che esso prevedeva un processo virtuoso di ricerca-azione e di formazione-azione che producesse ritorni anche dal basso.

Prima di lasciare alla lettura degli articoli, teniamo a ringraziare il collega John Humbley e le colleghe Alida Maria Silletti e Silvia Domenica Zollo dell'eccellente lavoro di curatela del presente volume, nonché per la riflessione comune che ha consentito di realizzarlo nelle migliori condizioni possibili. Un ringraziamento va anche al prof. Umberto Morelli, coordinatore del progetto *Artificial Intelligence for European Integration* che ha dato sempre supporto a ogni iniziativa e ha saputo gestire al meglio ogni difficoltà, non ultima, quella della pandemia da COVID-19 che ha cominciato a infuriare proprio nel periodo in cui il progetto ha preso avvio.

Un ringraziamento sentito va anche a tutte le persone che in modo più o meno diretto hanno contribuito alle ricerche. Senza di loro questo primo volume non sarebbe stato possibile.

Soprattutto, ci auguriamo di poter continuare a nutrire grazie a loro la riflessione comune su una tematica, come quella dell'IA, che è diventata e diventerà sempre più parte del nostro quotidiano.

Bibliografia

- Aeles Joris (2020). *Rapprocher le Parlement européen des citoyens grâce à l'IA*.
<https://pulse.microsoft.com/fr-be/transform-fr-be/government-fr-be/fa1-rapprocher-le-parlement-europeen-des-citoyens-grace-a-lia/>
- Bartoletti Ivana (2020). *An Artificial Revolution. On Power, Politics and AI*. Edimburgo: Indigo.
- Commissione europea (2020). *Libro bianco sull'intelligenza artificiale — Un approccio europeo all'eccellenza e alla fiducia*. COM(2020)65 final.
- International Organization for Standardization — ISO (2009). *Norma 23185. Assessment and benchmarking of terminological resources*. Ginevra: ISO.
- Le Cun Yann (2019). *Quand la machine apprend*. Parigi: Odile Jacob.
- Marzi Eleonora (2021). “La traduction automatique neuronale et le biais de genre : le cas des noms de métiers entre l'italien et le français”. *Synergies Italie*, 17, 19-36. <https://gerflint.fr/Base/Italie17/marzi.pdf>
- Organizzazione delle Nazioni Unite (2017). *Change. Rapport annuel 2017*. Ginevra: Nazioni Unite.

Introduction

Rachele Raus

This special issue of *De Europa* is the outcome of an interdisciplinary multilingual reflection carried out on research into linguistic rights, multilingualism and language varieties in Europe in the age of artificial intelligence¹. It is part of the *Artificial Intelligence for European Integration* project² I have been coordinating, headed by Umberto Morelli, with the backing of the Centre of European Studies To-EU of the University of Turin³ and co-financed by the European Commission.

Our aim was to investigate more generally the negative and/or positive outcomes of AI on language varieties and multilingualism, the latter a key value for the EU. The research started on 6 October 2020 with the paper given by Laurent Romary⁴, in charge of the Technical committee 37 of the International Standards Organisation (ISO) at the first international conference organised remotely by the University of Turin by the pilot group. Then, on 23 April 2021, a second remote conference was organized focusing on the specific topic we are dealing with. The papers given at this conference have been uploaded onto the project website⁵. The follow-up took the form of a morning workshop where those taking part, from several Italian and French universities, presented original research carried out on the didactics of foreign languages with regard to AI.

The discussion which ensued was so rich that it was decided to select some of the papers from the 23 and 24 April, to revise them in order to

Rachele Raus, Università di Bologna, rachele.raus@unibo.it

¹ To date this research initiative has involved around one hundred researchers from several EU countries, and, in addition, from Canada and Brazil.

² See <http://www.jmcoe.unito.it/home> (last accessed for all sites quoted 20 October 2021).

³ <https://www.dcps.unito.it/do/home.pl/View?doc=toeu.html>

⁴ <https://media.unito.it/?content=9806>

⁵ <http://www.jmcoe.unito.it/>

take into account all the meetings of the various working groups and people concerned and to add the contributions of those colleagues who had not been able to take part in the event, thereby further enriching an already plentiful debate. The result is a volume of original unpublished research being made generally available for the first time, at least in Europe.

1. How the issue emerged

As from the 1990s, several international organizations, UNESCO in particular⁶, started to question the ‘digital fracture’, caused, amongst other factors, by certain technologies being more widely available in some countries rather than others. Other similar dangers have emerged, for example the fact that languages that are spoken rather than written are less likely to be represented on the web, that large-scale corpora and big data are increasingly the property of private enterprise⁷, that some countries such as Canada or, in Europe, France and Germany, have invested massively in producing pages for the web and more generally in language industries, that most of the linguistic data and corpora are usually available only in English; that multilingual computer tools often use English as the pivot language (see Vetere’s chapter in the present issue), etc.

Working groups⁸ which we have organized around these questions indicate the following avenues of inquiry:

1. The question of language rights in relation to the possibility some countries have to invest in the language industries compared to some others. This question, raised by both linguists and lawyers in the working groups, was focused on by the United Nations special Rapporteur on Minority Issues, Fernand de Varennes⁹, at the 24 April 2021 conference. The debate will continue on the occasion of a conference organized by the Catholic University of Lille in the spring of 2022.
2. The need for careful selection of sources and data used to train AI algorithms.

⁶ See <http://www.unesco.org/new/index.php?id=50219>

⁷ Concerning document management for AI, see computer science specialists’ talk at the conference on 6 October 2020 at <https://media.unito.it/?content=9806>

⁸ The working groups were divided up by common language (FR, IT, EN).

⁹ https://drive.google.com/file/d/1YPSFX-v9ITB4N1tj_eFOI7Sinc9NBgcY/view

This second question raised other closely related issues, first of all the relations that these data have with forms of discrimination, termed ‘algorithmic discrimination’ by Bartoletti (2020) and that the relationship between these data and sources of ‘authority’ (*i.e.* authoritative source), which was raised in discussions with Laurent Romany during the 6 October conference. The present situation is that the authoritative source criterion proposed by ISO (2009 : 7) recommends that this source should be an official one or that it should guarantee criteria of quality, which indirectly implies that an institutional source is indeed required. This is probably one of the reasons that all the first training sessions for neural algorithms used in machine translation software or digital devices were performed with multilingual corpora sourced from international organisations. Another is that these institutions have a large quantity of multilingual data which are already available and pre-digested for automatic processing. That is what happened to the Google translator¹⁰. Its neural deep learning networks were trained on United Nations corpora (UN 2017: 120), and Facebook used multilingual documents from the European Parliament to train its automatic translators (Le Cun 2019: 277), or the Linguee concordancer¹¹ — its corpora constitute the training material for the automatic translator DeepL¹², a corpus which comes from EU institutions, UN, the Canadian federal government... That being said, using highly varied corpora and data which are not from institutions and therefore not necessarily checked and validated for quality has the effect of improving the overall performance of IT tools, compared with others using institutional corpora.

One topic that came up in the October conference is the case of the agreement of nouns and adjectives in the feminine. This tends to become particularly problematical in the case of nouns referring to occupations or professions when texts are drafted or translated automatically using tools based on deep learning neural networks trained on institutional corpora, which in turn rely on legal and/or political texts. The ‘neutral’ masculine then becomes the default option to refer to categories of people. This produces errors in the agreement in the feminine when texts are automatically translated into Romance languages, as is the case for Google or DeepL

¹⁰ <https://translate.google.com>

¹¹ <https://www.linguee.com>

¹² <https://www.deepl.com/en/translator>

translators, the most frequently used tools by Italian academics (see the contributions in the second part of the issue).

This problem could be solved, at least partially¹³, by using different corpora which are not necessarily validated and thus not institutional but which should be closer to actual usage and everyday language. The discussion about the data and the criteria used to establish them turned to the possibility of providing criteria before training neural networks. Thus the human intervention would be transferred from the current practice of revising the texts produced by the machine (i.e. after translation) to a preliminary step, where the linguistic and discursive implementation is used to train AI — selecting specific sub-corpora for specific tasks, providing criteria to improve supervised machine learning, etc.

These criteria would facilitate implementing the recommendations in the 2020 European white paper on human intervention in designing a tool to eliminate linguistic and gender discrimination and to impose ‘operational constraints’ in favour of multilingualism and linguistic variation (European Commission 2020: 19 and 21).

This is probably a time-consuming solution, that requires previous planning on what the ‘authority’ is to do in relation to AI training, also involving the task of tagging a large corpus to correspond to the needs of learning. That being said, it may be a time-consuming task but an indispensable one if AI is to fulfil its promise of promoting multilingualism. It is not enough for a machine translation unit to be multilingual if for example it uses English as the pivot language, as is the case for most of the tools produced by the current language industries.

In addition to the issue of language rights and data, the working groups on multilingualism and AI focused on the pedagogical implications, in particular how work in the classroom on languages and AI can be used for learning activities with tools of machine translation based on deep learning neural networks (see the second part of this issue).

¹³ Although there are gaps in feminine forms for certain professions and occupations in several Romance languages, notably French and Italian, this tendency is exacerbated in institutional corpora used in training algorithms which regularly resort to the masculine as a ‘neutral’ form to refer to persons in legal and thence to administrative discourse (see also Marzi 2021).

2. From a critical appraisal to AI promoting multilingualism

Although some contributors may be critical of the consequences of AI on multilingualism (see in particular the articles by Agresti and de Vecchi in this issue) and on a market, such as the language industries market, which today gives a biased view on language as mere vehicles of communication, or even interchangeable ‘codes’, it is by challenging AI that various ways of using these resources differently in favour of multilingualism can be achieved.

This was an invitation to view the data in a fresh light and to look into ways and means of designing and producing a multilingual application which would facilitate inclusive communication aimed at public service administrations. This application would be designed initially for Italian and then for other European languages (French, then German and finally Spanish and/or English, taking diatopic variation of these languages into account). It would be made with the support of the Polytechnical University of Turin based on criteria designed to develop AI promoting language variation and multilingualism. It is hoped to present the Italian application during the workshop to be organized at the end of 2023, where other similar initiatives will also be on show, such as the new automatic Facebook translator inaugurated in 2020, which no longer uses English as the pivot language¹⁴, or Microsoft AI Azure transcription used by the European Parliament, which takes into account different regional or national accents (Aeles 2020).

This workshop is intended to demonstrate how AI performance can be enhanced if only the right discursive and linguistic choices are made to take on board language variation, be it diaphasic, diatopic or diastratic.

As for the learning section, the working groups were set up during the 2020-2021 academic year and the results of the surveys are presented in the second part of this issue. These surveys have brought out how a specific training strategy can be used to heighten critical awareness of AI, which in turn results in a heightened awareness of the use of AI-based Language Industry tools.

This critical approach in no way undermines the initial confidence that the students say they have in AI, on the contrary it tends to reinforce confidence thanks to a better understanding of how these tools work and what can be made of them.

¹⁴The development team is coordinated by Prof. Tania Cerquitelli, whom we warmly thank.

In the long run, this issue of *De Europa* aims at promoting the improvement of AI on the one hand by implementing precise linguistic and discursive criteria before the learning takes place and on the other hand by increasing awareness of how certain criteria are favoured in the output. This way, the articles in this volume also intend to contribute to the final research outcomes of the project, planned for 2023, in particular:

- Recommendations for European decision makers on developing policies that favour artificial intelligence which respects multilingualism;
- Guidelines for those who work in vocational areas (technology, IT, standardization organizations, ...), which include suggestions for designing AI algorithms to promote multilingualism and diatopic and diastratic variation of languages, starting with the official languages of the EU.

3. Presentation of the contributions

This issue is made up of articles in one of three languages, Italian, French and English. They present the most salient features of the research carried out by the working groups over the first year and together make up the first volume of a trilogy, each volume presenting the results of the year's work. The approach is a very broad one so that highly specialized information can reach as many people as possible without compromising the scientific approach. A deliberate policy has been to use (gender)inclusive formulations, though no pressure has been put on any of the contributors to adopt any specific language choice. This results in some chapters preferring feminization, using the median point, hyphens or brackets, and other chapters — the majority in fact — where the choice has been for the so-called 'neutral' masculine.

This issue is in two parts. The first section is a general overview on AI as it is related to multilingualism and language variation. It also contains a number of specific case studies in particular concerning machine translation and AI, which go to make up the last part of this section. The second part of the volume contains the results of various surveys made in a teaching and learning context carried out in a number of European universities and business schools which took part in the initiative, namely Kedge Business School, EDEH, EM-Normandie, the Montpellier Business

School and the *École supérieure de commerce de Clermont* (Clermont Business School), the Catholic University of Lille, the Catholic University (Sacro Cuore) of Milan, the University of Naples-Orientale, the University of Turin and the University of Verona. This is also where some articles and studies are carried out by university students (first or second year Masters or postgraduate students), since the project was conceived as a research-action.

Before inviting our readers to discover the contents of this special issue we would like to thank John Humbley, Alida Maria Silletti and Silvia Domenica Zollo for their tireless efforts and punctual revision and formatting of the articles and for their participation in all the thought which has gone into the project which has come to fruition. We would also like to thank Umberto Morelli, who coordinates the project *Artificial Intelligence for European Integration* and who has always managed to steer the working groups in the right direction, even when the conditions became difficult notably with the Covid-19 pandemic.

We must also thank all those who contributed directly or less directly in the debates on the various topics.

We hope that the debate will continue and become richer as it continues, since artificial intelligence has become an indispensable part of our daily lives.

References

Aeles Joris (2020). *Rapprocher le Parlement européen des citoyens grâce à l'IA*. <https://pulse.microsoft.com/fr-be/transform-fr-be/government-fr-be/fa1-rapprocher-le-parlement-europeen-des-citoyens-grace-a-lia/>

Bartoletti Ivana (2020). *An Artificial Revolution. On Power, Politics and AI*. Edinburgh: Indigo.

European Commission (2020). *White paper on Artificial Intelligence — A European approach to excellence and trust*. COM(2020) 65 final.

International Organization for Standardization — ISO (2009). *Standard 23185. Assessment and benchmarking of terminological resources*. Geneva: ISO.

Le Cun Yann (2019). *Quand la machine apprend*. Paris: Odile Jacob.

Marzi Eleonora (2021). “La traduction automatique neuronale et le biais de genre : le cas des noms de métiers entre l’italien et le français”. *Synergies Italie*, 17, 19-36. <https://gerflint.fr/Base/Italie17/marzi.pdf>

United Nations (2017). *Change. Rapport annuel 2017*. Geneva: United Nations.