



Politique sur l'utilisation de la traduction automatique dans le contexte académique

Juillet 2023

Introduction

Le plus important en bref

Cette politique sert de guide aux collaborateurs et collaboratrices et aux étudiant e s de la Haute école spécialisée bernoise pour l'utilisation de la traduction automatique à des fins de compréhension et de rédaction de textes dans le cadre des études, de l'enseignement, de la recherche et de l'administration.

Ces dernières années, la traduction automatique neuronale (TAN ; neural machine translation, NMT), basée sur l'intelligence artificielle, a fait des progrès étonnants. Cela se reflète dans l'utilisation très répandue de la TAN dans la société, y compris dans le contexte académique. La TAN, aujourd'hui, est en mesure de fournir une qualité suffisante pour satisfaire une multitude d'objectifs et encourager la communication multilingue. Néanmoins, cette technologie a ses limites et comporte des risques – notamment en matière de protection des données, de confidentialité et de respect de la vie privée – dont les utilisateurs et utilisatrices doivent être conscient e s.

Les médias donnent en outre l'impression que le produit de la TAN peut être repris tel quel. Cependant, même les systèmes de TAN les plus avancés à ce jour présentent des problèmes de précision, d'adéquation ou de pertinence sous de nombreux aspects. Mais le produit de la TA étant grammaticalement correct et les erreurs imprévisibles, il est difficile pour les utilisateurs et utilisatrices de repérer et de corriger les inexactitudes. Ce phénomène est connu sous le nom de « false fluency », qui implique une maîtrise imparfaite de la langue. Les mots rarement utilisés, les phrases longues ou complexes, la terminologie spécialisée et les formulations délibérément ambiguës destinées à exprimer la réserve ou la probabilité (connues sous le nom de hedging ou d'enclosures) représentent un défi particulier pour les systèmes de TA. Les stéréotypes – notamment culturels, raciaux et sexuels – contenus dans le produit de la TA demeurent un problème pour les systèmes de TA. Il en va de même de la qualité du produit de la TA pour les langues moins répandues. Enfin, les systèmes de TAN accessibles au public effectuent des traductions à l'échelle de la phrase, ce qui pose d'importants problèmes de cohésion et de logique dans le texte, notamment s'agissant de références pronominales peu claires ou de terminologie incohérente.

Bien que l'on puisse s'attendre à une amélioration constante de la TAN, les systèmes de TA actuels produisent encore une qualité insuffisante à bien des égards tant au niveau du sens (saisie précise du message), du contexte (adaptation des références au groupe cible) et du texte (appropriation du style ou de la structure). Cette politique sert de guide aux collaborateurs et collaboratrices et aux étudiant e s de la Haute école spécialisée bernoise pour l'utilisation de la traduction automatique dans le cadre des études, de l'enseignement, de la recherche et de l'administration.

Définitions

Le plus important en bref

Par traduction automatique (machine translation, MT ; en français TA), on entend la traduction d'un texte d'une langue à une autre à l'aide d'un logiciel, sans intervention humaine. Le produit de la TA peut être utilisé sous sa forme brute (non traitée) ou faire l'objet d'une post-édition (révision humaine appelée machine translation post-editing ou MTPE).

Traduction brute

Usage de la traduction automatique pour obtenir une compréhension globale, mais éventuellement imprécise, du message du texte source.

Communication à haut risque

Textes à forte visibilité (par exemple, les textes publiés à l'extérieur) qui sont destinés à susciter un comportement ou une réaction spécifique de la part du lectorat ou qui pourraient nuire à la réputation, entraîner des risques financiers ou juridiques s'ils étaient mal compris.

Communication à faible risque

Textes à faible visibilité (par exemple destinés à un usage exclusivement interne) ou textes utilisés occasionnellement et uniquement à des fins d'information.

Traduction automatique

Traduction d'un texte d'une langue à une autre à l'aide d'un logiciel, sans intervention humaine.

Traduction automatique neuronale

Dernière génération de systèmes de traduction automatique à partir de données ou de corpus et basés sur des réseaux neuronaux artificiels. Les systèmes de traduction automatique neuronale (TAN) sont capables de produire des résultats de haute qualité et sont entraînés à « apprendre » à traduire des textes et des mots qu'ils n'ont pas rencontrés pendant l'apprentissage.

Post-édition

Édition du produit de la TA pour corriger les erreurs de langue, de sens ou de style.

L'étape de post-édition peut être divisée en différentes catégories en fonction de la personne qui l'effectue, du texte et de son objectif, ainsi que du public cible :

- Post-édition par des utilisateurs et utilisatrices occasionnels
- Post-édition de communication à faible risque (voir définition ci-dessous) afin de corriger les erreurs facilement identifiables
- Post-édition par des traducteurs et traductrices qualifiés
- Post-édition légère de communication à haut risque (voir la définition ci-dessous), afin de s'assurer que le sens a été correctement transmis, sans correction stylistique
- Post-édition complète pour les communications à haut risque (voir la définition ci-dessous), afin de garantir une qualité comparable à celle d'une traduction humaine professionnelle

Pré-édition

Traitement d'un document avant sa traduction automatique. Cette étape consiste par exemple à corriger des erreurs dans le texte source, à simplifier les structures de phrases ou à supprimer les ambiguïtés.

Produit de TA brut

Produit de TA qui n'a pas été traité ou révisé par l'utilisateur ou l'utilisatrice.

Langue source

La langue à partir de laquelle un texte est traduit.

Texte source

Texte à traduire.

Langue cible

Langue dans laquelle un texte est traduit

Texte cible

Version traduite du texte source.

Traduction

Processus de transfert du message et du sens d'un texte de la langue source à la langue cible.

Protection des données

Le plus important en bref

Si vous utilisez un système de TA gratuit (par exemple Google Translate, DeepL), ne saisissez pas de données personnelles ou d'informations confidentielles (y compris les données issues des activités de recherche internes). Lorsque vous utilisez un système de TA payant (p. ex. DeepL Pro), vous devez vérifier la politique de confidentialité de l'entreprise en question. En effet, certaines entreprises assurent la protection de l'ensemble des données, ce qui permet la transmission de données personnelles. Indépendamment de cela, vous êtes tenu-e-s de respecter le règlement général sur la protection des données (RGPD), notamment en ce qui concerne les données à caractère personnel et les données éventuellement soumises à un devoir de confidentialité ou de discrétion. Ne saisissez jamais de données à caractère personnel sensibles.

a. Systèmes de TA gratuits (p. ex. Google Translate, DeepL)

Aucun système de TA n'est gratuit. Les utilisateurs et utilisatrices paient avec leurs textes. Les données injectées peuvent être réutilisées comme données d'entraînement pour les algorithmes ou revendues à d'autres sociétés. C'est pourquoi il est illicite de saisir des données personnelles ou des informations confidentielles (y compris les données issues des activités de recherche internes) lorsque vous utilisez un système de TA gratuit.

b. Système de TA payant (p. ex. DeepL Pro)

Certains prestataires de services de TA proposent un abonnement payant, dans le cadre duquel il est garanti que les textes des utilisateurs et utilisatrices ne seront pas réutilisés. Certaines entreprises peuvent même assurer la protection de l'ensemble des données afin de permettre la transmission de données personnelles. Toutefois, l'utilisation de ces systèmes doit respecter le règlement général sur la protection des données, notamment en ce qui concerne les données à caractère personnel et les données éventuellement soumises à un devoir de confidentialité ou de discrétion. Par conséquent, lorsque vous utilisez un système de TA, vous devez vérifier la politique de confidentialité de l'entreprise en question.

Introduire les informations suivantes dans un système de TA est, dans tous les cas, proscrit :

- informations confidentielles telles que définies par la structure de classification des informations de la BFH ;
- informations couvertes par le secret professionnel (notamment les données relatives aux patients ou à la santé) ;
- informations soumises contractuellement au devoir de discrétion ; données personnelles particulièrement sensibles (données relatives à la santé, à la vie privée ou à l'appartenance ethnique ; données faisant référence à l'appartenance ou aux activités religieuses, politiques ou syndicales ; données relatives aux poursuites et sanctions administratives ou pénales et données relatives aux mesures d'aide sociale ou à l'internement forcé).

Transparence

Le plus important en bref

Si vous utilisez la traduction automatique, vous devez toujours indiquer dans une clause de non-responsabilité (disclaimer) si le texte produit résulte d'une traduction automatique brute ou retravaillée (post-éditée).

En principe, les lecteurs et lectrices d'un texte doivent toujours être informé e s de l'utilisation de la TA. Par exemple, si un courriel est rédigé à l'aide d'un système de TA, il convient d'ajouter une clause de non-responsabilité indiquant si le courriel est le produit de la TA, brut ou post-édité. Les destinataires du message peuvent de cette façon se faire une meilleure idée de la qualité de la traduction et guetter d'éventuelles erreurs. Vous trouverez ci-dessous quelques exemples de clauses de non-responsabilité que vous pouvez adapter en fonction du niveau de post-édition.

Produits de TA bruts ou légèrement post-édités par un utilisateur ou une utilisatrice occasionnel-le :

« Ce texte a fait l'objet d'une traduction automatique en [insérer la langue cible] et peut comporter des erreurs. Vous trouverez ci-dessous le texte original en [insérer la langue source]. »

Post-édition minimale par un traducteur ou une traductrice qualifié-e (également appelée « post-édition légère ou rapide ») :

« Ce texte a été traduit automatiquement et légèrement retravaillé afin d'en garantir la lisibilité et l'exactitude. La version [insérer la langue cible] est fournie à titre d'information uniquement. »

Si l'étape de post-édition complète est effectuée par un traducteur ou une traductrice qualifié e, aucune clause de non-responsabilité n'est requise.

Remarques sur l'utilisation de la TA

Le plus important en bref

Lorsque vous utilisez la TA, vous devez toujours respecter les règles applicables en matière de protection des données ou de confidentialité. En outre, les principes suivants s'appliquent :

a. Utilisation de la TA dans le cadre des études et de l'enseignement

Si vous faites traduire par un système de TA un texte que vous n'avez pas rédigé vous-même, et que vous utilisez le texte produit par la TA sans mentionner la source originale, il s'agit d'un plagiat. Les étudiant e s doivent respecter les consignes données par leurs enseignant e s pour l'utilisation de la TA dans le cadre de leurs travaux et de leurs évaluations.

b. Utilisation de la TA dans la recherche et dans l'administration

Avant de faire traduire un texte par un système de TA, il convient de vérifier si une erreur éventuellement générée par le logiciel pourrait constituer une infraction punissable ou porter atteinte à la réputation, entraîner des préjudices juridiques, des dommages financiers ou de sérieux malentendus s'accompagnant de conséquences graves. Pour toute communication sensible, il est impératif de faire appel à un traducteur ou à une traductrice qualifié e. Lisez la politique complète pour en savoir plus sur les types de textes qui conviennent à la TA.

En cas d'utilisation de la TA, les règles en matière de droits d'auteur ou d'autres questions d'ordre éthique doivent être respectées. Les principes suivants s'appliquent à cet égard :

a. Utilisation de la TA dans le cadre des études et de l'enseignement

Si vous faites traduire par un système de TA un texte que vous n'avez pas rédigé vous-même et que vous utilisez le texte produit par la TA sans mentionner la source originale, il s'agit d'un plagiat.

- Les étudiant e s et les enseignant e s sont tenu e s de signaler l'utilisation de produits de TA bruts ou post-édités.
- Les étudiant e s doivent respecter toutes les consignes données par leurs enseignant e s pour l'utilisation de la TA dans le cadre de leurs travaux et évaluations (y compris les examens, les travaux de semestre, les travaux de bachelor et de master et autres).

b. Utilisation de la TA dans la recherche et dans l'administration

L'utilisation de la TA dépend du contexte d'utilisation du texte, de son objectif et du public cible. Si les textes sont destinés à une utilisation interne pour une période limitée ou pour une communication informelle, la présence de quelques imprécisions dans la traduction ne devrait pas avoir de conséquences graves. Toutefois, si les textes sont utilisés de manière permanente ou pour une communication externe formelle, des toute erreur ou imprécision peut nuire à la réputation.

Pour les textes (non publiés) qui ne sont utilisés que pendant une période limitée ou pour une communication interne à faible risque, il est possible d'utiliser le produit brut ou post-édité de la TA, à condition de :

- joindre une clause de non-responsabilité indiquant que le texte a été traduit par une machine ;
- faire suivre la traduction par le texte source, afin que les lecteurs et lectrices puissent comparer le produit de la TA avec l'original en cas d'ambiguïté, d'incertitude, etc. ; et enfin
- respecter les points de la liste de contrôle TA relative aux étapes de pré-édition et de post-édition.

Les courriels, les présentations internes, les textes non publiés ou les rapports internes constituent des exemples de textes qui ne sont utilisés que pendant une période limitée ou pour une communication interne à faible risque.

Les textes permanents (publiés) ou utilisés pour une communication externe à haut risque doivent être retravaillés ou traduits par un traducteur ou une traductrice qualifié-e. Si le texte a fait l'objet d'une post-édition légère ou rapide, cela doit être communiqué au moyen d'une clause de non-responsabilité. En cas de post-édition complète, aucune clause de non-responsabilité n'est requise.

Parmi les exemples de textes utilisés de manière permanente ou pour une communication externe à haut risque, on peut citer les attestations officielles, les évaluations de performance, les textes académiques ou hautement spécialisés tels que les publications scientifiques, les textes créatifs nécessitant une adaptation en fonction du contexte culturel tel que le contenu des sites internet, les rapports annuels et le matériel de marketing.

Aide et conseils

Le plus important en bref

Pour toute question relative à cette directive, adressez-vous au service spécialisé Bilinguisme et plurilinguisme à l'adresse bilinguisme@bfh.ch.

Le produit de la TA peut varier considérablement en fonction du type de texte, de la paire de langues et du système de TA. Si vous n'êtes pas sûr e qu'un texte se prête à une traduction automatique ou qu'il requière une étape de post-édition professionnelle, contactez le service de traduction de la BFH, de votre département ou un traducteur ou une traductrice qualifié-e.

Pour toute question relative à cette politique, adressez-vous au service spécialisé Bilinguisme et plurilinguisme à l'adresse bilinguisme@bfh.ch.