

Mieux comprendre  
les outils d'aide

Délégation générale à la langue française et aux langues de France

à la traduction

Grâce au développement des technologies de la langue, les outils d'aide à la traduction font aujourd'hui partie intégrante de l'environnement de travail des traducteurs. Des progrès constants, issus de travaux de recherche, font évoluer les solutions et les rendent de plus en plus performantes. Cependant, ces outils ne sont réellement efficaces que s'ils sont véritablement adaptés au travail du traducteur et intégrés dans les processus de traduction.

La délégation générale à la langue française et aux langues de France (ministère de la Culture et de la Communication) a souhaité mesurer l'adéquation entre les outils d'aide à la traduction et les besoins quotidiens des traducteurs.

Ce document présente les résultats de la consultation publique menée auprès des traducteurs en septembre et octobre 2014, et contient des informations pratiques sur les différents types d'outils qui leur sont proposés. Les informations relatives à ces outils, jusqu'à présent inégales et dispersées, ont été regroupées et synthétisées dans un travail inédit qui permet d'une part de cerner les pratiques et habitudes de la communauté, d'autre part de mieux appréhender certains outils et familles d'outils de traduction à travers leurs différentes fonctionnalités.

Cet instantané, tout en apportant des informations utiles aux différents acteurs de la traduction, permettra, je l'espère, de mesurer les progrès accomplis depuis quelques années et d'éclairer les choix de politique publique en matière de soutien à la traduction. Sa cible est multiple : les utilisateurs finaux, les porteurs de projets (institutions, organismes de recherche) et les éditeurs de produits et solutions.

À travers ces résultats, nous comprenons mieux les priorités à définir, comme la nécessité d'encourager la formation des traducteurs à l'usage des outils d'aide à la traduction. Les technologies de la langue représentent en effet une opportunité importante que le monde de la traduction ne doit pas manquer.

Loïc Depecker



Délégué général à la langue française et aux langues de France

# Remerciements

Ce guide a été réalisé grâce au concours de nombreux services et organismes publics et privés ; la délégation générale à la langue française et aux langues de France (DGLFLF) entretient avec eux des relations étroites.

Nous remercions également l'ensemble des personnes qui ont participé aux entretiens, et notamment les membres du Groupe Interministériel pour la Traduction (GIT) organisé conjointement par le service de traduction du ministère des Finances et des Comptes publics et la délégation générale à la langue française et aux langues de France, ainsi que les nombreux professionnels qui ont répondu à la consultation publique.

L'ensemble de cette étude a été réalisée au moyen des logiciels libres LimeSurvey et LibreOffice par Alain Couillault, à la demande de la délégation générale à la langue française et aux langues de France. Les données statistiques complètes relatives à cette étude sont disponibles sous forme tabulée à l'adresse suivante :

<http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Langue-francaise-et-langues-de-France/Politiques-de-la-langue/Langues-et-numerique/Les-technologies-de-la-langue-et-la-normalisation/Mieux-comprendre-les-outils-d-aide-a-la-traduction>

Ce document est librement mis à disposition

sous les conditions de la licence Creative Commons CC-BY-SA 3.0.



<http://creativecommons.org/licenses/by-sa/3.0/fr/>

# Avertissement

Nous attirons votre attention sur le fait que les données présentées dans ce document sont issues de la consultation publique sur les outils d'aide à la traduction menée par la délégation générale à la langue française et aux langues de France entre septembre et octobre 2014.

Compte tenu de l'évolution rapide de l'offre (nouveaux produits, mises à niveau des produits existants), ce document pourra faire l'objet d'actualisations directement en ligne sur le site internet de la délégation : [www.culturecommunication.gouv.fr/Politiquesministerielles/Langue-francaise-et-langues-de-France](http://www.culturecommunication.gouv.fr/Politiquesministerielles/Langue-francaise-et-langues-de-France)

Pour nous signaler de nouveaux logiciels et outils qui ne figurent pas dans la présente édition ou pour toute autre remarque ou suggestion, vous pouvez nous contacter par courrier électronique à : [outilstraduction@culture.gouv.fr](mailto:outilstraduction@culture.gouv.fr)

# Sommaire

<b>1. À propos de ce document</b>	<b>6</b>
1.1 La consultation publique	6
1.2 Méthodologie employée	7
1.3 Comment lire les tableaux	9
1.4 Limites de la consultation	11
<b>2 Reconnaissance Optique de Caractères et Conversion de documents</b>	<b>15</b>
2.1 Comment ça marche?	15
2.2 Les avis des traducteurs	19
2.3 Astuces et commentaires de traducteurs	21
<b>3 Outils à mémoire de traduction</b>	<b>22</b>
3.1 Comment ça marche?	22
3.2 Les avis des traducteurs	27
3.3 Astuces et commentaires des traducteurs	36
3.4 Commentaires sur les produits	39
<b>4 Traduction automatique</b>	<b>41</b>
4.1 Comment ça marche?	41
4.2 Les avis des traducteurs	45
4.3 Astuces et commentaires de traducteurs	47
<b>5 Gestion de terminologie</b>	<b>49</b>
5.1 Comment ça marche?	49
5.2 Les avis des traducteurs	51
5.3 Astuces et commentaires de traducteurs	54
<b>6 Extraction terminologique</b>	<b>55</b>
6.1 Comment ça marche?	55
6.2 Les avis des traducteurs	57

# 1. À propos de ce document

Les outils d'aide à la traduction deviennent progressivement indispensables dans le métier de traducteur, et peuvent accompagner le travail de traduction à plusieurs étapes du processus. Nous rappelons ici que **le rôle de ces outils est d'aider le traducteur, mais jamais de le remplacer**. Il est encore aujourd'hui et probablement pour longtemps inconcevable d'imaginer remplacer un travail de traduction humain par des routines logicielles.

Notons également que si ces outils trouvent naturellement leur place pour la **traduction de documents professionnels**, ils ne sont pas adaptés au travail de traduction littéraire, qui relève d'un processus d'écriture pouvant difficilement être assisté par des outils informatiques. La question de l'outillage éventuel de la traduction littéraire n'entre donc pas dans le périmètre de cette étude.

En outre, l'utilisation d'outils de traduction impacte fortement **le métier de traducteur** : de plus en plus, le traducteur qui utilise des outils d'aide à la traduction va être amené à faire un travail de **post-édition**, c'est-à-dire qu'il va intervenir après le passage d'un logiciel pour faire des choix terminologiques, syntaxiques ou grammaticaux, valider des propositions, ou bien encore corriger des traductions qui auraient été faites automatiquement.

Au-delà de la question du choix des outils et de leurs fonctionnalités, cela implique **des changements majeurs dans le processus de travail du traducteur**, qui ne font pas l'objet de ce document. Les implications de ces changements sur la question du coût et du mode de facturation de la traduction ou encore des rémunérations des traducteurs ont également été exclues du périmètre de cette étude. Ces questions sensibles mériteraient à elle seules de faire l'objet d'une réflexion approfondie.

## 1.1 La consultation publique

Entre septembre et octobre 2014, la délégation générale à la langue française et aux langues de France (DGLFLF) – ministère de la Culture et de la Communication, a conduit une **consultation publique** auprès de traducteurs francophones pour recueillir leurs avis sur les outils d'aide à la traduction qu'ils utilisent.

Ce document présente de manière synthétique les données recueillies lors de cette consultation auprès des répondants. Sans prétendre être un guide d'achat, un comparatif exhaustif des solutions disponibles sur le marché ou encore une étude statistique ou scientifique sur la question, il a pour objectif d'apporter des informations **aux traducteurs et aux services de traduction** leur permettant de mieux comprendre l'utilité et les fonctionnalités des différents outils d'aide à la traduction disponibles sur le marché. Il sera aussi, indirectement, utile aux éditeurs de produits et solutions, qui trouveront dans ce document la somme des avis et évaluations de nombreux utilisateurs.

Conscient des limites de cet exercice (voir. Partie 1.4 « Limites de la consultation »), nous estimons que cette publication présente **un intérêt pédagogique** pour toutes les personnes ou équipes de traduction désireuses de s'équiper en outils et qui ne disposent que de peu d'informations à ce sujet. En effet, la première constatation faite lors de l'élaboration du questionnaire puis du présent document est que l'information disponible sur les outils d'aide à la traduction est insuffisante et à ce jour très parcellaire.

Nous espérons donc que le fruit de ce travail constitue non pas un aboutissement mais plutôt **un point de départ** vers une meilleure information sur les outils d'aide à la traduction, déterminants pour le métier de traducteur, et appelons de nos vœux une initiative coordonnée dans ce sens de la part des différents acteurs et spécialistes du monde de la traduction.

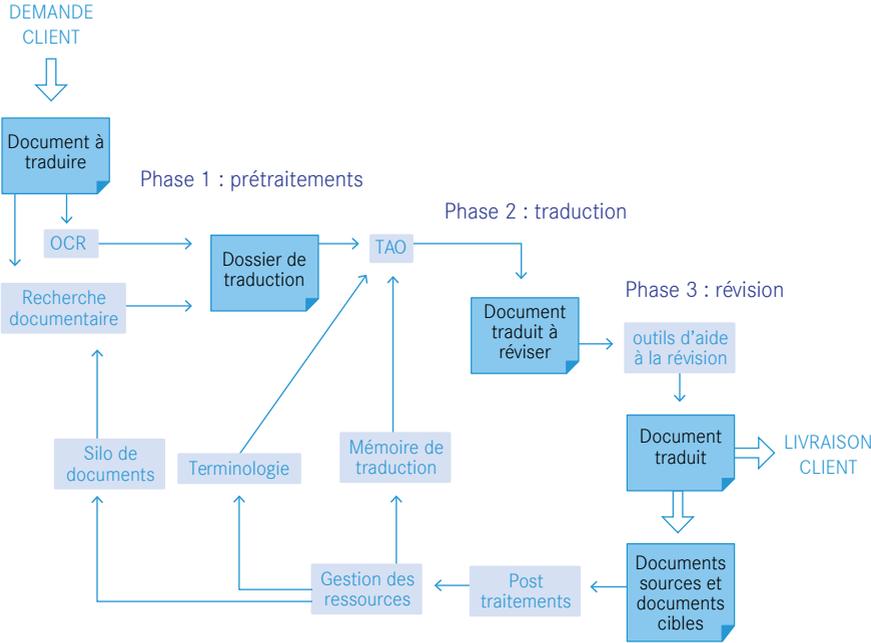
## 1.2 Méthodologie employée

Cette enquête s'est déroulée sous la forme d'un **questionnaire en ligne** diffusé largement, via les associations, les établissements de formation et les différents réseaux qui animent cette communauté. Ce document propose une synthèse des réponses des nombreux traducteurs – plusieurs centaines – qui y ont participé.

**Composé d'environ 120 questions**, dont une bonne part était optionnelle, le questionnaire a été élaboré par un groupe de travail composé d'experts, d'enseignants et de chercheurs dans le domaine de la traduction, en collaboration avec des traducteurs professionnels.

Cette première étape a permis d'identifier **un processus de traduction type** décrit dans le schéma ci-dessous : le document à traduire (la demande client) fait d'abord l'objet d'un pré-traitement (conversion, recherche documentaire), il est ensuite traduit, éventuellement avec l'aide d'outils de Traduction Assistée par Ordinateur (TAO), puis corrigé avant d'être remis au client. L'utilisation d'outils nécessite que des gestionnaires maintiennent différentes bases (de documents, de terminologies, de mémoires de traductions). Tous les outils décrits dans ce processus ne sont pas nécessairement mis en œuvre.

Illustration 1 : Processus typique de traduction



Dans ce processus ont été identifiées **cinq familles d'outils** : la conversion de documents et la reconnaissance optique de caractères<sup>1</sup>, les mémoires de traduction, la traduction automatique, la gestion terminologique et l'extraction terminologique.

<sup>1</sup> Ou OCR pour *Optical Character Recognition* dans sa version anglophone.

Ce document est organisé en cinq parties correspondant à chacune de ces familles d'outils. Chaque partie contient une présentation générale des outils et des techniques qu'ils mettent en œuvre, puis des résultats statistiques, et enfin, une synthèse des avis et astuces.

Les moyennes présentées dans le document sont directement tirées des réponses recueillies lors de la consultation et n'ont pas fait l'objet de vérifications.

Par ailleurs, les commentaires des répondants figurant dans ce document ont été reproduits, car ils semblaient utiles pour les lecteurs. Certains commentaires, positifs ou négatifs, citant explicitement certains produits disponibles sur le marché n'ont en revanche pas été repris dans ce rapport.

## 1.3 Comment lire les tableaux

La synthèse statistique des réponses fournit trois types d'informations : un indice de satisfaction, un indice d'utilité et un taux de réponses positives. Les questions qui étaient posées sont indiquées en légende des tableaux.

### 1.3.1 Indice de satisfaction

L'indice de satisfaction est un nombre compris entre 0 et 5, 5 étant la meilleure note.

Il est calculé, selon les questions posées, soit en faisant la moyenne des notes de 1 à 5 attribuées par les traducteurs, soit à partir du nombre de personnes interrogées indiquant que la fonctionnalité répond effectivement au besoin du traducteur.

### 1.3.2 Indice d'utilité

L'indice d'utilité est un nombre compris entre 0 et 5, 5 étant la meilleure note.

Il est calculé à partir du nombre de traducteurs déclarant que la fonctionnalité est utile ou non et donne ainsi une indication sur l'utilité de la fonctionnalité telle qu'elle est perçue par les traducteurs.

## Illustration 2 : Indices d'utilité et de satisfaction

	Utilité	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Alignement automatique des mémoires de traduction	4.3	4.4	2.5	2.5	1.5	1.3	2.5	1.6	1.2	1.6
Édition de la mémoire	4.2	4.3	5.0	5.0	4.5	5.0	5.0	4.8	5.0	5.0
Extraction automatique de la terminologie	1.9	0.0	-	-	3.0	3.3	2.5	2.5	2.0	0.0
Import de glossaires	4.2	2.9	5.0	5.0	4.5	5.0	5.0	4.0	4.7	4.4
Mise à disposition des mémoires de traductions à utilisateur extérieur	3.2	4.0	0.0	0.0	4.7	2.5	3.3	4.7	4.4	4.3
Export de mémoires de traduction	4.4	4.4	0.0	0.0	4.8	2.5	3.3	4.8	4.7	4.0

### 1.3.3 Taux de réponses positives

Le taux de réponses positives est un nombre situé entre 0 et 5.

Il était demandé aux traducteurs si une fonctionnalité était proposée ou non par l'outil. Les résultats sont présentés sous la forme d'une note allant de 1 à 5 correspondant à la proportion de réponses positives. Une note de 5 (sur 5) indique que toutes les personnes ont répondu par l'affirmative, et une note de 2,5 (sur 5), que la moitié des répondants a répondu par l'affirmative.

Des réponses intermédiaires (entre 1 et 4) peuvent surprendre s'agissant pour l'essentiel de savoir si une fonction est proposée ou non pour un outil. Ces notes intermédiaires peuvent être l'indice de plusieurs situations, dont voici une liste non exhaustive observée dans cette enquête :

- la fonction est inconnue ou méconnue de la personne qui a répondu ;

- une fonction similaire mais légèrement différente est proposée par l'outil ;
- une nouvelle version est diffusée par l'éditeur qui justement propose cette fonctionnalité.

Notre démarche visait à recueillir les avis des professionnels, nous avons donc décidé de restituer ces valeurs intermédiaires en l'état, pour indiquer que ce point nécessite une attention particulière.

Par exemple :

Illustration 3 : Exemples de réponses à une question binaire

	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Recherche d'une chaîne de caractères	6	10	10	10	10	10	10	10	10
Recherche documentaire (ET, OU, troncature)	2	10	10	4	2	5	2	3	2
Recherche par expressions régulières	5	-	-	5	1	5	4	4	2

## 1.4 Limites de la consultation

Un certain nombre de limites sont apparues pendant et après la consultation publique, et doivent être prises en compte pour la bonne compréhension de ce document.

### 1.4.1 Sur la fiabilité des résultats

La fiabilité des données présentées dans ce document est tout d'abord liée à la **quantité de répondants** pour chaque série de questions ou d'outils. Le nombre de répondants, pour chaque catégorie d'outils, a ainsi été indiqué dans chaque rubrique de ce document pour permettre de mieux apprécier la quantité de données recueillies par type d'outil.

En outre, la fiabilité des résultats doit également prendre en compte le **niveau de connaissance des répondants**, qui n'a pas fait l'objet de vérifications particulières. Ainsi, un répondant pourra répondre « oui » à la question « cette fonctionnalité est-elle proposée par le produit ? » parce qu'il pense que la fonctionnalité est disponible, alors qu'en réalité, elle ne l'est pas.

Ces erreurs d'appréciation peuvent aussi être expliquées par la formulation de la question qui pouvait parfois prêter à confusion. C'est particulièrement le cas pour les résultats dont la moyenne obtenue est médiane (proche de 2,5) alors que la question était de type binaire : « cette fonctionnalité est-elle proposée ? ».

Nous rappelons que les données présentes dans le document n'ont fait l'objet d'**aucune « expertise » particulière**, s'agissant de données brutes collectées lors de la consultation en ligne. Sur la base de ce premier document, il apparaît utile de faire vérifier par un expert la présence ou non de telle ou telle fonctionnalité dans les différents outils étudiés, notamment dans les cas où les réponses des traducteurs font apparaître un doute (note médiane).

## 1.4.2 Sur le nombre et les versions des solutions proposées

Lors de l'élaboration du questionnaire s'est posée la question d'intégrer **les différentes versions de chaque logiciel d'aide à la traduction**. Compte-tenu du grand nombre de versions disponibles pour chaque produit, il a malheureusement été impossible de distinguer, dans la consultation, les différentes versions disponibles pour chaque produit. En effet, d'une part il aurait été difficile de présenter les résultats de manière lisible et synthétique (trop grand nombre de colonnes), d'autre part la fiabilité des résultats aurait été mauvaise du fait d'un trop grand éparpillement des réponses sur les différents outils et leurs versions respectives.

**Des réserves** doivent donc être émises sur la validité des résultats obtenus pour certains logiciels qui ont fortement évolué au fil des versions proposées à leurs clients, ou qui sont disponibles suivant plusieurs variantes comprenant des fonctionnalités différentes. C'est notamment le cas pour « Trados », les résultats de cette consultation combinant des évaluations pour toutes les versions de cet outil dans ses variantes « Trados », « Trados SDLX » ou encore « Trados Studio ». Nous attirons donc particulièrement l'attention des lecteurs sur ce point.

### 1.4.3 Sur la représentativité des répondants.

La consultation publique a permis de collecter plus de 700 réponses, la plupart ne couvrant pas l'ensemble des outils évalués. Néanmoins, nous rappelons que cette démarche n'est ni scientifique ni statistique : en aucun cas il n'a été constitué d'échantillon représentatif et les données ont été recueillies [sur la base du volontariat](#).

Par ailleurs, bien qu'un dispositif empêchant de répondre plusieurs fois au questionnaire sur le site internet ait été prévu, ce dispositif est facile à contourner (en utilisant plusieurs postes ou appareils connectés par exemple).

Enfin, il faut noter que la consultation, qui était d'une durée limitée (deux mois), n'a peut-être pas laissé le temps à chaque répondant de compléter ses réponses ou bien plus simplement, de participer.

### 1.4.4 Sur les catégories d'outils

Il est apparu au cours de l'élaboration du questionnaire que la catégorisation par famille d'outils en termes de besoins fonctionnels pouvait se révéler délicate. Certains outils ont ainsi été regroupés au sein de la même famille, mais présentent des modalités d'accès ou de fonctionnement différentes, [ce qui rend difficile leur comparaison chiffrée](#).

C'est le cas notamment des outils à mémoire de traduction (partie 3), dans laquelle figurent des outils disposant d'une interface propre et des outils pour lesquels le traducteur travaille directement dans son éditeur de texte.

C'est aussi le cas pour les outils de traduction automatique, qui comprennent des outils en ligne, des outils commerciaux, des moteurs nécessitant un développement informatique.

### 1.4.5 Sur les questions abordées

Plusieurs aspects des outils de traduction n'ont pas été couverts par cette consultation publique, et gagneraient à être traités dans un travail ultérieur sur le même sujet :

- **La question du prix et du mode de licence** des produits a été volontairement écartée, l'objectif étant de se concentrer sur l'aspect fonctionnel des outils et non d'évaluer le rapport qualité/prix de telle ou telle solution. Les produits libres et gratuits ont été évalués de la même manière que les produits onéreux.
- **La question de la sécurité des données** n'a pas été évaluée dans le cadre de la consultation. Il s'agit pourtant d'une problématique très forte pour les services de traduction, qui se pose dès lors qu'il y a des échanges de données de traduction avec des tiers via internet, encore plus lorsque l'outil proposé est entièrement en ligne sur internet (pas d'application installée sur le poste de l'utilisateur). Les données traduites, bien que parcellaires, peuvent être critiques pour une entreprise soumise à la concurrence internationale par exemple, et il est recommandé aux utilisateurs de ces outils de faire preuve de la plus grande vigilance quant à la transmission à l'extérieur de ces données.
- Enfin, **la question de l'export des mémoires de traduction** (évaluée dans la figure 14 du présent rapport) et des différents formats proposés par les outils à mémoire de traduction mériterait d'être plus largement détaillée, cette question étant particulièrement importante dans un contexte d'échanges de plus en plus massifs des mémoires entre structures de traduction.

#### 1.4.6 Sur l'actualisation des données

Les éditeurs de logiciels faisant évoluer et progresser leurs outils constamment, il est donc probable que les lacunes qui semblaient apparaître pour tel ou tel produit lors de la consultation publique en septembre-octobre 2014 soient comblées, corrigées ou améliorées prochainement (si ce n'est déjà fait).

Par conséquent, il serait utile de disposer d'**un observatoire permanent des outils de traduction** dans lequel figureraient des données actualisées et expertisées sur les fonctionnalités des outils les plus utilisés par les traducteurs. Une telle initiative serait profitable à tous les métiers de la traduction et permettrait de choisir efficacement l'outil le plus adapté à ses besoins mais dépasse le périmètre de cette étude.



est, pour certains logiciels, corrigé ensuite automatiquement par une analyse linguistique sommaire du document pour la correction d'orthographe ou le repérage d'erreurs systématiques.

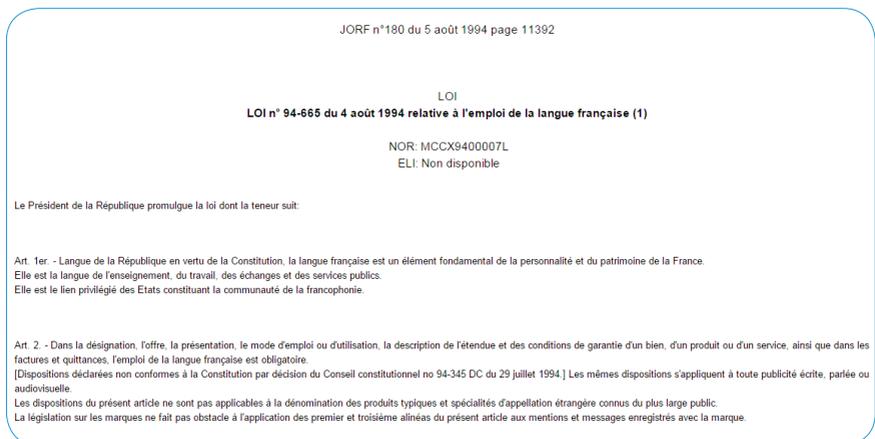
### 2.1.2 Conversion de documents PDF texte

Le format PDF (Portable Document Format) vise à garantir l'affichage d'un document quel que soit le support, écran ou papier, et quel que soit le type d'appareil utilisé : ordinateur, téléphone, tablette. Les pages d'un document au format PDF sont décrites par un langage informatique qui assure cette unicité d'affichage. Ce langage décrit tous les éléments de mise en forme (police, grasse, images...) mais ne décrit pas la structure des pages (styles, hiérarchie des titres, renvois...). Les outils de conversion visent donc à analyser le code informatique du fichier PDF, à en extraire le texte et à en déduire les informations de structure.

### 2.1.3 Illustration

Ces deux situations sont illustrées par les exemples ci-dessous. L'illustration 3 montre le texte d'origine sous forme d'image.

Illustration 4 : Image du texte de la loi n° 94-665 du 4 août 1994



### 2.1.3.1 Reconnaissance de caractères

L'application d'un outil de reconnaissance de caractères produit le texte ci-dessous :

JORF n° 180 du 5 août 1994 page 11392

LOI

LOI n° 94-665 du 4 août 1994 relative à l'emploi de la langue française [1]

NOR: MCCCX9400007L

ELI: Non disponible

Le Président de la République promulgue la loi dont la teneur suit :

**Art. 1<sup>er</sup>.** - Langue de la République en vertu de la Constitution, la langue française est un élément fondamental de la personnalité et du patrimoine de la France.

Elle est la langue de l'enseignement, du travail, des échanges et des services publics.

Elle est le lien privilégié des Etats constituant la communauté de la francophonie.

**Art. 2.** - Dans la désignation, l'offre, la présentation, le mode d'emploi ou d'utilisation, la description de l'étendue et des conditions de garantie d'un bien, d'un produit ou d'un service, ainsi que dans les factures et quittances, l'emploi de la langue française est obligatoire.

[Dispositions déclarées non conformes à la Constitution par décision du Conseil constitutionnel no 94-345 DC du 29 juillet 1994.] Les mêmes dispositions s'appliquent à toute publicité écrite, parlée ou audiovisuelle.

Les dispositions du présent article ne sont pas applicables à la dénomination des produits typiques et spécialités d'appellation étrangère connus du plus large public.

La législation sur les marques ne fait pas obstacle à l'application des premier et troisième alinéas du présent article aux mentions et messages enregistrés avec la marque.

### 2.1.3.2 Conversion de documents PDF texte

En partant du même texte, mais cette fois au format PDF texte, un outil de conversion de documents PDF texte fournit le texte suivant :

JORF n° 180 du 5 août 1994 page 11392

LOI

LOI n° 94-665 du 4 août 1994 relative à l'emploi de la langue française (1)

NOR: MCCX9400007L

ELI : Non disponible

Le Président de la République promulgue la loi dont la teneur suit :

**Art. 1<sup>er</sup>.** - Langue de la République en vertu de la Constitution, la langue française est un élément fondamental de la personnalité et du patrimoine de la France.

Elle est la langue de l'enseignement, du travail, des échanges et des services publics.

Elle est le lien privilégié des Etats constituant la communauté de la francophonie.

**Art. 2.** - Dans la désignation, l'offre, la présentation, le mode d'emploi ou d'utilisation, la description de l'étendue et des conditions de garantie d'un bien, d'un produit ou d'un service, ainsi que dans les factures et quittances, l'emploi de la langue française est obligatoire.

[Dispositions déclarées non conformes à la Constitution par décision du Conseil constitutionnel n°94345 DC du 29 juillet 1994.] Les mêmes dispositions s'appliquent à toute publicité écrite, parlée ou audiovisuelle.

Les dispositions du présent article ne sont pas applicables à la dénomination des produits typiques et spécialités d'appellation étrangère connus du plus large public.

La législation sur les marques ne fait pas obstacle à l'application des premier et troisième alinéas du présent article aux mentions et messages enregistrés avec la marque.

### 2.1.3.3 Choisir un logiciel d'OCR ou de conversion

Ces deux techniques (OCR et conversion) répondent au même besoin : transformer un document dans un format non-modifiable en un document de travail. La qualité des documents produits dépend principalement de la qualité et de la nature du document source : document image ou document PDF texte, texte dactylographié ou manuscrit.

L'OCR obtient actuellement un taux de reconnaissance des caractères compris généralement entre 90 et 98 %, ce qui signifie 2 à 10 erreurs sur 100 caractères<sup>2</sup> pour des documents dactylographiés. La reconnaissance d'écriture manuscrite est quant à elle fortement dépendante de l'écriture du rédacteur.

---

<sup>2</sup> Voir à ce sujet le chapitre « La numérisation des documents : principes et évaluation de la performance », *Les documents écrits de la numérisation à l'indexation par le contenu*, Paris, Lavoisier, 2006.

## 2.2 Les avis des traducteurs

**61 traducteurs** ont répondu aux questions de la consultation relatives à la famille d'outils « Reconnaissance optique de caractères et conversion de documents ». Ce chiffre relativement faible comparé aux autres familles d'outils signifie que la fiabilité des résultats obtenus est moins grande que pour les autres catégories d'outils traités dans ce document. Néanmoins, **55 %** des répondants déclarent utiliser ces outils, ce qui montre que leur usage est relativement courant.

Le choix d'un logiciel d'OCR et de conversion ne dépend pas uniquement de sa capacité à restituer le texte ou les styles du document source. Lorsqu'elle est nécessaire, cette étape est généralement la toute première du processus, et parfois le précède même lorsqu'elle vise à effectuer un comptage de mots pour établir un devis à moindre frais. Pour d'autres usages (alimenter une mémoire de traduction, extraire de la terminologie, ou utiliser un logiciel de traduction automatique ou assistée) une édition a posteriori ou post-édition est presque toujours nécessaire.

Pendant, ces outils restent peu intégrés dans les solutions d'aide à la traduction disponibles sur le marché, et les traducteurs ou les gestionnaires souhaitent pouvoir intervenir en cours de numérisation pour corriger des fautes récurrentes, gérer des gabarits de numérisation ou déclencher des traitements par lot.

### 2.2.1 Les logiciels

La consultation publique concernait trois outils d'OCR et conversion :

Produits	Abby Fine Reader	Adobe Acrobat	Nuance Omnipage
Éditeur	ABBYY	Adobe	Nuance
Pays	États-Unis	États-Unis	États-Unis
Site	<a href="http://www.abbyy.com">www.abbyy.com</a>	<a href="http://www.adobe.com">www.adobe.com</a>	<a href="http://www.nuance.com">www.nuance.com</a>

### 2.2.2 Tableaux de synthèse

Le tableau 1 propose une synthèse des avis des traducteurs concernant la qualité des documents produits par l'OCR, le tableau 2, les fonctions annexes, et le tableau 3, l'ergonomie et le support.

## 2.2.3 Fonctions essentielles

Tableau 1 : Évaluez la qualité du document produit par l'OCR en choisissant une note de 1 à 5 (5 étant la meilleure note)

	Abby Fine Reader	Adobe Acrobat	Nuance OmniPage
Reconnaissance texte caractères latins	<b>3.9</b>	3.8	4
Reconnaissance texte caractères arabes	<b>2.4</b>	2	1.7
Reconnaissance texte caractères cyrilliques	<b>2.8</b>	2	1.7
Reconnaissance texte idéogrammes	<b>2.1</b>	2	1.5
Moyenne	<b>2.8</b>	2.5	2.2

## 2.2.4 Fonctions annexes

Tableau 2 : Concernant les fonctions suivantes :

		Abby Fine Reader	Adobe Acrobat	Nuance OmniPage
	Utilité	Avis	Avis	Avis
Correction du texte généré	4	3.3	1.7	<b>3.6</b>
Reconnaissance automatique de la langue dans laquelle le document est écrit	4	<b>5</b>	4.7	<b>5</b>
Traitement des documents par lot	2.5	<b>4</b>	2.5	3
Création de gabarits de reconnaissance pour le traitement de documents similaires	1	2	0	<b>5</b>
Le logiciel OCR produit-il un rapport d'erreurs indiquant clairement les zones qui peuvent être mal reconnues?		2	1.6	<b>2.1</b>
Moyenne		3.3	2.1	<b>3.7</b>

## 2.2.5 Ergonomie et support

Tableau 3 : En ce qui concerne le logiciel OCR, choisissez une note de 1 à 5 (5 étant la meilleure note) pour évaluer :

	Abby Fine Reader	Adobe Acrobat	Nuance Omnipage
Utilisabilité <sup>3</sup>	3	2.6	2.8
Ergonomie	3.5	3	3.6
Support	3.2	2.2	1.7
Moyenne	2.9	2.4	2.6

## 2.3 Astuces et commentaires de traducteurs<sup>4</sup>

- *Sauf lorsqu'ils sont numérisés, les documents PDF sont généralement produits par un traitement de texte à partir d'un document original que le donneur d'ordre peut bien souvent fournir.*
- Le coût des outils peut être important par rapport au gain qu'ils apportent. On peut cependant, en fonction de ses besoins, se tourner vers des logiciels libres (AbleWord, FreeOCR, Tesseract par exemple) ou des sites Internet qui proposent la conversion en ligne.
- Le texte restitué par ces outils peut nécessiter un travail important de correction, en particulier quand la mise en page est un peu compliquée : tableaux, en-tête et pieds de pages, présences d'images. Les caractères accentués ou les césures restent mal reconnus par certains logiciels. Il est ainsi parfois préférable de n'extraire que le texte et de refaire l'ensemble de la mise en page.

<sup>3</sup> La question posée était : « Pour servir de base à la traduction, le document converti avec le logiciel OCR est : le texte produit est exploitable sans modification, exploitable après quelques modifications/éditions, difficilement exploitable, inutilisable ».

<sup>4</sup> Cette section est rédigée à partir des commentaires laissés par les traducteurs et ne saurait en aucun cas être considérée comme un avis de la Délégation concernant la qualité des outils mentionnés.



Les algorithmes d'alignement se fondent généralement sur deux types de critères, formels ou linguistiques, pour découper les textes en segments, puis les appairer. Dans l'exemple ci-dessous, le fait que *Découvrir le parapente* soit un titre est un bon indicateur à la fois de découpage et d'appariement : le segment en anglais correspondant (*Discover paragliding*) est également un titre, les sens des mots sont proches. D'autres indices permettent également de découper le texte en segments, comme les différentes ponctuations (dans l'exemple ci-dessous, le point final suivi d'une espace et d'une majuscule, et le point d'exclamation sont des marques relativement fiables de fins de phrases). Le découpage peut être plus fin et reconnaître, grâce à des techniques d'analyse linguistiques, des segments de plus petites longueurs (par exemple, des propositions ou des syntagmes nominaux). Évidemment, plus un segment est court, plus il est probable qu'il en existe une traduction.

### Exemple :

#### **Découvrir le parapente**

Pour les aventuriers qui veulent encore plus, découvrez les joies du parapente !  
Volez haut dans les airs et admirez le paysage.

#### **Discover paragliding**

For the more adventurous, discover the joy of hang gliding! Fly high up in the air and enjoy the scenery.

*Texte 1 : Alignement de textes*

Cette opération de segmentation est effectuée pour les textes des deux langues, l'alignement consiste ensuite à comparer les propriétés des segments d'une langue à l'autre : les segments ont-ils les mêmes styles ? Contiennent-ils des mots traduits ? Sont-ils ordonnés ?

Cette opération produira par exemple le tableau suivant pour l'exemple ci-dessus :

Français	Anglais
Découvrir le parapente	Discover paragliding
Pour les aventuriers qui veulent encore plus, découvrez les joies du parapente !	For the more adventurous, discover the joy of hang gliding!
Volez haut dans les airs et admirez le paysage.	Fly high up in the air and enjoy the scenery.

Un découpage plus fin grâce à une analyse linguistique pourra donner le tableau de correspondance suivant :

Français	Anglais
Découvrir le parapente	Discover paragliding
Pour les aventuriers qui veulent encore plus,	For the more adventurous,
découvrez les joies du parapente !	discover the joy of hang gliding!
Volez haut dans les airs	Fly high up in the air
admirez le paysage.	enjoy the scenery.

La **comparaison** des segments évalue le nombre d'éléments communs pour fournir un indice, nul en cas de correspondance exacte, qui permet de classer les propositions faites au traducteur. La plupart des solutions du marché permettent d'indiquer un seuil au-delà duquel les segments trop dissemblables ne sont pas proposés. Cet indice peut être complété par la prise en compte du contexte. En effet, le logiciel pourra proposer en priorité un segment ayant un contexte proche de celui du texte proposé. Cela est particulièrement pertinent lorsque les segments comparés sont courts.

### 3.1.2 Choisir une solution de mémoire de traduction

Les logiciels de mémoire de traduction sont des alternatives à d'autres outils, tels que les logiciels de traduction automatique ou d'aide à la rédaction. Ils impliquent généralement une modification dans les habitudes de travail du traducteur qu'il faut prendre en compte afin de bénéficier au mieux des gains qu'ils apportent : cohérence des traductions et diminution des temps de traduction notamment.

Un logiciel de mémoire de traduction doit être configurable, afin de réduire le plus possible la présence de paires de segments indésirables dans la mémoire de traduction. En effet, plus les segments proposés au traducteur sont pertinents, plus le gain de temps est important. Pour la même raison, selon les répondants, on privilégiera les solutions qui donnent la possibilité de contrôler manuellement le contenu des mémoires. Cela peut être fait à plusieurs niveaux, en supprimant ou modifiant des segments, ou via la terminologie. La majeure partie du temps du traducteur sera consacrée à choisir, intégrer, modifier, compléter, mettre en forme... les textes proposés par le logiciel, et, probablement, à changer ses habitudes de travail.

Il est donc essentiel de s'assurer de la bonne ergonomie de l'outil choisi, et de sa bonne intégration dans l'espace de travail du traducteur. Enfin, une mémoire de traduction est une base de données, qui doit être maintenue et mise à jour, et peut nécessiter des outils dédiés, ainsi que des compétences particulières. Il faudra donc prendre en compte les éventuels besoins en formation, voire dédier des ressources spécifiquement à cette maintenance.

### 3.1.3 Les langues traduites

Le tableau 4 (taux de réponses langue source/langue cible) indique pour chacune des langues le pourcentage de répondants qui traduisent depuis (*langue source*) ou vers (*langue cible*) cette langue.

Plus d'un traducteur sur deux traduit depuis l'anglais. L'étude a été conduite en français et diffusée principalement à des services ou des formations francophones. Un traducteur sur deux travaille vers le français et, parmi eux, un sur deux travaille sur le couple français/anglais. Le nombre de réponses pour les autres langues ne permet pas de tirer des conclusions fines sur la pertinence des outils en fonction des langues.

Tableau 4 : taux de réponses langue source/langue cible

Langues	Anglais	Arabe	Espagnol	Français	Italien	Japonais	Mandarin	Néerlandais	Portugais	Russe	Autre
Source	51 %	1 %	15 %	15 %	7 %	-	1 %	2 %	3 %	1 %	4 %
Cible	1-	2 %	6 %	57 %	7 %	2 %	1 %	2 %	4 %	3 %	4 %

### 3.1.4 Les logiciels

La consultation publique concernait neuf outils de traduction assistée par ordinateur (TAO) :

Produits	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Éditeur	Across Systems GmbH	Systran International	Systran International	Kilgray Translation technologies	MultiCorpora	Lingua et Machina	SDL Language Technologies	Wordfast LLC	Atril
Pays	Allemagne	Corée du Sud	Corée du Sud	Hongrie	Canada	France	Royaume-Uni	États-Unis	États-Unis
Site	www.across.net	www.systransoft.com	www.systransoft.com	www.memoq.com	www.multicorpora.com	www.similis.org	www.translation-zone.com	www.wordfast.net	www.atril.com

## 3.2 Les avis des traducteurs

**172 traducteurs** ont répondu aux questions de la consultation relatives à la famille d'outil « à mémoire de traduction ». Les logiciels de mémoire de traduction semblent désormais fréquemment utilisés par les traducteurs. En effet, **75%** des répondants déclarent utiliser au moins un outil à mémoire de traduction dans le cadre de son activité.

Notons que pour 100 traducteurs (sur la base des réponses de la consultation), 20 connaissent ou utilisent deux outils différents, plus de cinq d'entre eux en connaissent ou utilisent trois, et six en utilisent quatre ou plus.

Le tableau ci-dessous précise, pour chaque type de structure, le pourcentage de traducteurs qui disent utiliser chacun des outils. La solution Trados vient en tête du nombre de réponses : une sur deux pour les formations, les structures publiques, et les indépendants ; une sur trois pour les structures privées. MemoQ vient en second pour le nombre de solutions installées. Cet outil est utilisé par un tiers des répondants travaillant dans des organisations internationales. Plus de la moitié des traducteurs indépendants qui ont répondu utilisent Trados, alors que plus d'un quart des structures publiques utilisent Multitrans.

Tableau 5 : taux de réponse par solutions et par type de structure

	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Formation	-	-	-	11%	5%	5%	53%	11%	16%
Structures publiques	-	-	-	7%	27%	7%	4%	13%	7%
Structures privées	13%	4%	4%	8%	8%	8%	33%	4%	17%
Structures internationales	-	-	-	33%	-	-	67%	-	-
Agences de traduction	1%	-	-	2%	2%	2%	49%	1%	6%
Traducteurs indépendants	4%	-	-	2%	1%	-	51%	17%	7%
Total du tableau	6%	-	-	17%	4%	2%	49%	13%	8%

## 3.2.1 Ergonomie

Un traducteur passe une grande partie de son temps devant son ordinateur, l'ergonomie de son poste de travail est donc importante, ce qui semble être bien pris en compte par les éditeurs de logiciels, les scores pour ces aspects allant de 4,5 à 4,9. Les fonctions par nature plus collaboratives, comme la possibilité de poser des commentaires, ne sont pas perçues comme des priorités par les traducteurs.

Tableau 6 : Score et utilité : Pour la solution de TAO que vous utilisez, quelle est votre opinion concernant les fonctionnalités suivantes liées à l'ergonomie ?

	Utilité	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Possibilité de créer des macros spécialisée	4.6	4.4	-	-	4.6	1.7	3.3	4.5	<b>4.7</b>	4.1
Saisie automatique (auto-complétion)		4.2	-	-	4.4	5	5	4.4	4.5	4.5
Propagation des corrections	4.5	4.4	-	-	<b>4.8</b>	2.5	3.3	4.5	4.7	4.2
Gestion des styles et des tableaux	4.6	4.4	-	-	4.6	1.7	3.3	4.5	<b>4.7</b>	4.1
Possibilité de poser des commentaires	3.2	<b>5</b>	<b>5</b>	<b>5</b>	4.8	2.5	<b>5</b>	4.7	3.8	<b>5</b>
Compatibilité de l'outil avec d'autres logiciels (notamment de traitement de texte)	4.9	<b>5</b>	<b>5</b>	<b>5</b>	4.7	3.8	<b>5</b>	4.7	<b>5</b>	4.6
Note globale d'ergonomie		3.3	2.9	3	3	<b>3.4</b>	1.8	3	3.1	3.3
Moyenne		<b>4.4</b>	4.3	4.3	<b>4.4</b>	2.9	3.8	4.3	<b>4.4</b>	4.3

### 3.2.2 Accès à la mémoire

Une mémoire de traduction doit sans cesse être mise à jour. Les traducteurs souhaitent pouvoir la modifier en cours de travail (indice d'utilité 4,8), et voir ces modifications propagées automatiquement sur le texte à traduire. Là encore, les fonctions collaboratives (partage de mémoire, travail simultané sur un projet) ne sont pas privilégiées, probablement parce qu'elles ne sont nécessaires que pour traduire des documents volumineux, avec des échéances courtes et qui donc nécessitent l'intervention de plusieurs traducteurs.

La majorité des traducteurs en formation et de ceux qui travaillent dans des organismes publics ou internationaux considère que la « possibilité de travailler simultanément à plusieurs sur un même texte » est un besoin pour eux. C'est également le cas d'un traducteur en agence sur deux. En outre, si un traducteur indépendant sur deux considère qu'il n'a pas besoin de cette possibilité, plus d'un sur cinq déclare que le logiciel qu'il utilise n'est pas conforme à son besoin pour cette fonctionnalité. Il n'y a cependant pas de corrélation entre le nombre de pages traduites annuellement dans un service et l'utilisation de fonctions collaboratives.

Tableau 7 : Pour la solution de TAO que vous utilisez, quelle est votre opinion concernant les fonctionnalités suivantes de mise à jour de la mémoire temporaire de traduction ?

	Utilité	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Modification de la mémoire de traduction ou de la terminologie en cours de traduction	4.8	5	5	5	5	5	5	4.9	5	5
Possibilité de gérer simultanément des projets de traduction associés à des mémoires différentes	4.2	5	5	5	5	5	5	4.7	4.7	5
Possibilité d'alimenter simultanément une mémoire temporaire liée à un projet de traduction	3.8	4.3	5	5	4.5	2.5	5	4.5	3.9	3.9
Possibilité de travailler simultanément sur un même texte	2.8	5	5	5	3.8	2.5	5	4.1	2.9	3.8
Moyenne		4.8	5	5	4.6	3.8	5	4.6	4.1	4.4

### 3.2.3 Fonctions de recherche

La mémoire de traduction, au-delà des correspondances entre langues, fournit une base riche d'exemples dans les langues cible et source. Tous les outils proposent donc des fonctions de recherche de termes : recherche par simple chaîne de caractères, recherche documentaire (et, ou, troncatures...), ou recherche par expression régulière.

Tableau 8 : Quelles fonctions de recherche sont fournies pour explorer la mémoire de traduction ? (scores ramenés pour 10 traducteurs).

	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Recherche d'une chaîne de caractères	3	5	5	5	5	5	5	5	5
Recherche documentaire (ET, OU, troncature)	1	5	5	2	1	2.5	1	1.5	1
Recherche par expressions régulières	2.5	-	-	2.5	0.5	2.5	2	2	1
Moyenne	2.5	5	5	3.2	2.2	3.3	2.7	2.8	2.3

Tableau 9 : Les résultats de la recherche donnent-ils accès au contexte du segment retrouvé, si oui, quel contexte est affiché ?

	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Contexte affiché: phrase	3	-	-	4.5	2.5	5	4	5	3.5
Contexte affiché: paragraphe	2	-	-	1	1.5	2.5	1.5	1.5	2.5
Contexte affiché: document entier	-	-	-	0.5	4	2.5	0	0.5	0.5
Moyenne	2.5			2	2.7	3.3	1.8	2.3	2.2

### 3.2.4 Intégration des segments de mémoire

Les textes présents dans la mémoire de traduction diffèrent parfois de ceux du document à traduire par quelques éléments *variables* (dates, montant, données chiffrées). Ces segments doivent être insérés dans le texte cible avec des mises en formes éventuellement différentes selon la langue. Par exemple l'expression 9 200,50€ en français devra être traduite par €9,200.50 en anglais et par 9.200,50 EUR<sup>5</sup> en allemand.

Tableau 10: Concernant les correspondances indiquées comme exactes incluant des éléments variables (dates, montants, chiffres)

	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Ces éléments variables sont clairement identifiés dans l'interface	4.5	5	5	5	3.5	5	5	4.5	5
Ces éléments variables sont correctement intégrés dans le texte cible	5	-	-	4.5	-	5	4.5	3.5	5
Moyenne	4.8	5	5	4.8	3.5	2	4.8	4	5

### 3.2.5 Compatibilité

L'utilisation des outils sous licence libre reste limitée au profit d'outils propriétaires tels que Microsoft Office, ou des éditeurs intégrés à l'outil de traduction. Pratiquement aucun des logiciels n'est perçu comme étant compatible avec Libre Office. Ces données correspondent effectivement à la présence de ces produits en entreprise: selon une étude de Forrester<sup>6</sup>, en 2011, seulement 13% des personnes interrogées déclaraient utiliser des solutions alternatives libres, et ce chiffre est tombé à 5% en 2013.

<sup>5</sup> Source <http://j.poitou.free.fr/pro/html/typ/allemand.html>

<sup>6</sup> Market Update: Office 2013 And Productivity Suite Alternatives, Philipp Karcher Peter Burris, TJ Keitt, Andrew Smith, octobre 2013.

Tableau 11 : La solution de TAO est-elle utilisable à partir des éditeurs de texte du marché suivants ?

	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Microsoft Word	2	5	5	2	3	2.5	2.5	3.5	1.5
Libre Office	0.5	0	0	1	0	0	0.5	0.5	0.5
La solution intègre son propre éditeur	3	5	5	3	1	4	2.5	2	3

### 3.2.6 Qualité de l'éditeur intégré

Pour les fonctionnalités essentielles, l'éditeur intégré dans la solution de traduction assistée répond globalement aux besoins des traducteurs selon eux.

Tableau 12 : Score des fonctions de l'éditeur intégré

	Utilité	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Gestion des styles	4.7	5	5	5	5	5	5	4.5	5	5
Gestion de la typographie	4.6	5	5	5	5	5	5	4.5	5	5
Gestion des tableaux	4.4	5	5	5	5	5	5	4.6	5	5
Accès à la mémoire de traduction	4.6	5	5	5	5	5	5	4.8	5	5
		5	5	5	5	5	5	4.6	5	5

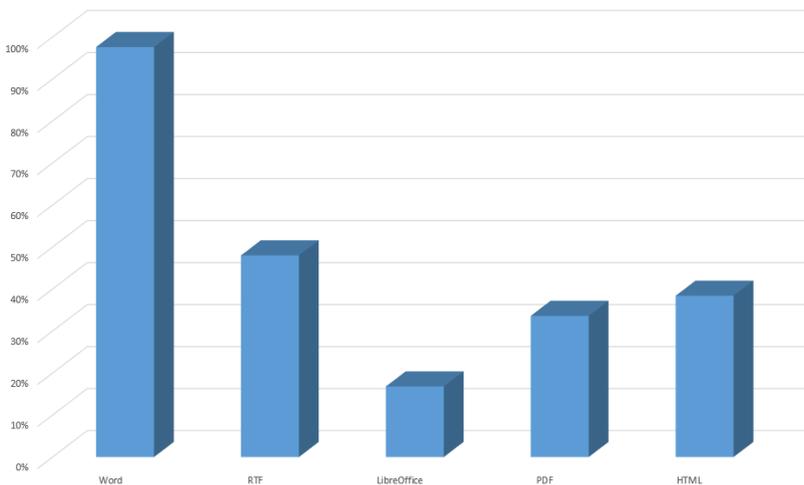
## 3.2.7 Mémoire de traduction : les fonctions d'administration

La qualité et le gain de temps apportés par un outil d'aide à mémoire de traduction dépend pour une grande part de la bonne administration de cette mémoire, qui nécessite des compétences particulières. Le temps et les compétences qui doivent être consacrés à cette administration doivent être pris en compte au moment du choix et de la mise en place d'un outil.

### 3.2.7.1 Format de documents

Les documents à partir desquels les administrateurs construisent ou nourrissent la mémoire de traduction sont majoritairement au format *Word*. Notons que la part de documents au format PDF reste importante (34%), alors qu'un tiers des personnes interrogées déclarent que les documents issus des logiciels d'OCR ou de conversion doivent être corrigés avant de pouvoir être utilisés (voir tableau 2 page 16).

Illustration 5 : Quels formats de fichiers utilisez-vous pour nourrir la mémoire de traduction ?



### 3.2.7.2 Taux de reprise

Le *taux de reprise* est le pourcentage de segments d'un texte à traduire déjà présents (à l'identique, ou approchant) dans la mémoire de traduction. Ce taux est un bon indicateur de l'effort de traduction, et donc de son coût. Notons que pour ce tableau, le nombre de scores compris entre deux et trois est important.

Tableau 13 : La solution comprend-elle des fonctions pour évaluer...

	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Le taux de reprise	3	<b>5</b>	<b>5</b>	2	1.5	1.5	2	2.5	3
Le nombre de segments (concordance floue) présents dans la base	2.5	<b>5</b>	<b>5</b>	2.5	2	2.5	2.5	3	2.5
Le nombre de segments (concordance exacte) présents dans la base	2.5	-	-	2.5	2	1.5	2.5	<b>3</b>	2
Moyenne	2.8	<b>3.5</b>	<b>3.5</b>	2.3	1.8	1.8	2.3	2.8	2.5

### 3.2.8 Gestion de la mémoire de traduction

Tableau 14 : Quelle est votre opinion concernant les fonctionnalités suivantes ?

	Utilité	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Alignement automatique des mémoires de traduction	4.3	<b>4.4</b>	2.5	2.5	1.5	1.3	2.5	1.6	1.2	1.6
Édition de la mémoire	4.2	4.3	5	5	4.5	<b>5</b>	<b>5</b>	4.8	<b>5</b>	<b>5</b>
Extraction automatique de la terminologie	1.9	0.0	-	-	3	<b>3.3</b>	2.5	2.5	2	0.0

Import de glossaires	4.2	2.9	5	5	4.5	5	5	4	4.7	4.4
Mise à disposition des mémoires de traductions à un utilisateur extérieur	3.2	4	0.0	0.0	4.7	2.5	3.3	4.7	4.4	4.3
Export de mémoires de traduction	4.4	4.4	0.0	0.0	4.8	2.5	3.3	4.8	4.7	4
Import de mémoires de traduction existantes	4.7	3.9	0.0	0.0	5	2.5	3.8	4.9	5	4.5
Partage d'une partie de la mémoire de traduction	3.2	5	5	5	3.8	5	5	5	3.6	5
Possibilité de modifier la concordance des segments	4	2	0	0	2	0	0	4	2	0
Moyenne	3.8	3.4	2.2	2.2	3.8	3	3.4	4	3.6	3.2

### 3.2.9 Autres fonctions

Tableau 15: Qualité du découpage et formats

	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Qualité du découpage	4.4	5	5	4.6	5	5	4.5	4.3	5
Liste des formats d'import suffisante	5	5	5	4.4	5	5	4.1	4.4	4.4
Moyenne	4.7	5	5	4.5	5	5	4.3	4.4	4.7

## 3.2.10 Ergonomie et support pour les administrateurs

Tableau 16 : Ergonomie et support

	Across	Business Translator	Enterprise server	MemoQ	MultiTrans	Similis	Trados	Wordfast	Déjavu
Ergonomie de l'interface du traducteur	2.4	3	3	2.5	1.7	<b>4.3</b>	2.4	2.7	2.3
Ergonomie de l'interface de l'administrateur	1.8	3	3	1.3	0.7	<b>3.3</b>	1.4	1.1	1.4
Service support	1.6	3	3	1.8	0.9	<b>2.8</b>	1.4	1.9	1.9
Moyenne	1.9	3	3	1.9	1.1	<b>3.5</b>	1.7	1.9	1.9

## 3.3 Astuces et commentaires des traducteurs

L'utilisation quotidienne de ces outils permet d'apprécier les avantages et les inconvénients des fonctionnalités qu'ils proposent. Les traducteurs ont partagé leurs impressions dans le questionnaire et ont mis en avant certains points qu'il convient d'examiner lorsque l'on choisit un outil plutôt qu'un autre.

### 3.3.1 Ergonomie

La possibilité de traduire directement dans le document source, avec la contrainte d'avoir à récupérer les segments dans un fichier bilingue est appréciée différemment selon les traducteurs.

Lors du déploiement d'une nouvelle version de logiciel, il arrive que l'interface, même si elle est de meilleure qualité, nécessite une nouvelle prise en main et donc un temps d'adaptation voire de formation qu'il ne faut pas négliger.

La visualisation du contexte des segments similaires de la mémoire de traduction, la possibilité de mettre en vis-à-vis deux textes similaires pour les comparer, la possibilité d'élargir la fenêtre dans laquelle on travaille, et l'affichage de concordances plus

étendues dans le concordancier, ainsi qu'une meilleure reconnaissance des styles et des niveaux des textes ont été mentionnés comme étant des fonctionnalités auxquelles les traducteurs sont sensibles.

La reconnaissance vocale peut être utile pour pouvoir dicter une traduction plus rapidement.

Paradoxalement, la traduction française de certaines interfaces du logiciel est parfois très approximative.

### **3.3.2 Administration de la mémoire de traduction**

L'ergonomie de la gestion de la mémoire de traduction varie grandement d'un outil à l'autre, et peut nécessiter des compétences particulières selon l'outil ou les documents à traduire. La possibilité de mettre à jour des mémoires de traduction par import de contenus par lot, ou d'attribuer des méta-données aux mémoires sont des fonctions utiles à l'administrateur. En outre, malgré l'existence de standards dans ce domaine, les mémoires de traduction sont parfois peu compatibles avec les autres outils. Des fonctions de nettoyage de la mémoire de traduction sont nécessaires pour la bonne maintenance des bases. L'endroit où est stockée la mémoire de traduction (sur le poste ou en ligne) peut avoir des conséquences sur l'environnement de travail, notamment pour les traducteurs nomades.

### **3.3.3 Liens vers des sources externes**

Disposer dans l'interface de liens directs vers des ressources terminologiques externes (IATE ou dictionnaires techniques de type Ernst) ou d'autres outils, par exemple de traduction automatique, est un plus.

### **3.3.4 Fonctions de recherche**

Une part non négligeable du travail de traduction consiste à rechercher de l'information : dans une base terminologique, dans une liste de commentaires, ou bien de rechercher des segments comparables dans la langue cible ou dans la langue source.

Ces recherches doivent être ergonomiques, disponibles immédiatement et efficaces, particulièrement concernant la façon dont la recherche est exprimée et celle dont les résultats sont affichés.

### **3.3.5 Import et export**

L'import de documents PDF est de qualité satisfaisante, mais l'export des documents au format PDF se fait en texte.

Les fonctions d'import et d'export de documents cibles ou sources (PDF, documents bureautiques) sont de qualités variables selon l'outil et devraient être testées sur les documents proches de ceux que l'on aura à traduire si l'on souhaite mettre en œuvre de tels outils. Par exemple, certains outils fonctionnent très bien avec des traitements de textes, mais sont inefficaces pour des documents issus de tableaux ou d'outils de création de présentation.

### **3.3.6 Support et suivi**

De nombreux traducteurs soulignent que la qualité du support n'est pas toujours à la hauteur du prix du service ou de la licence annuelle. Les autres sources d'information ne sont par ailleurs pas toujours satisfaisantes : documentation lacunaire ou service d'assistance technique peu réactif. Ces situations peuvent conduire à des propositions de prestations facturées, il convient donc de vérifier l'étendue de l'offre comprise dans le contrat de support proposé par l'éditeur.

### **3.3.7 Gestion de projets**

Les gestionnaires de projets voudraient que soient incluses des fonctions de comptabilité, avec, notamment, des indications de la durée d'utilisation (chronomètre) pour évaluer la rentabilité du travail, et des fonctions de gestion de projet, comme la communication avec le client de messagerie électronique.

## 3.4 Commentaires sur les produits<sup>7</sup>

Globalement, les traducteurs réclament plus de transparence et de pouvoir sur la mémoire de traduction (possibilité de l'actualiser, de modifier les alignements de segments en temps réel), ainsi que la diversification de ses formats d'import et d'export.

Certains outils sont appréciés pour leurs possibilités de formatage similaires à celles des logiciels de traitement de texte habituels, ou pour une utilisation facile avec des fichiers tabulés ou des présentations, avec des résultats néanmoins un peu moins fiables. Dans ce cas, les répondants suggèrent le recours aux versions professionnelles des outils, lorsqu'elles existent.

Certains outils riches en fonctionnalités peuvent être difficiles à appréhender sans formation ou aide du support, ce qui peut être problématique pour les traducteurs indépendants notamment. Cette richesse fonctionnelle, parfois au détriment de l'ergonomie, peut donc s'avérer être un frein à l'adoption de l'outil, ou bien ne pas être pleinement exploitée.

En ce qui concerne les questions d'ergonomie et de prise en main, les réponses sont variables selon les outils : certains semblent satisfaisants voire faciles d'utilisation, alors que d'autres présentent des prises en main complexes.

Les répondants notent qu'il est possible d'utiliser des outils externes pour remplacer les fonctions de nettoyage et de mise à jour des mémoires et des glossaires (tableurs, éditeurs de textes) parfois manquantes.

Enfin, les répondants apprécient les outils pouvant être reliés à des bases de données, ce qui facilite les recherches documentaires.

### 3.4.1 Autres outils

D'autres logiciels à mémoire de traduction, ou visant à apporter une aide au traducteur, sont commercialisés sous différentes formes : sous licences payantes, gratuits, logiciels à la demande, voire en production participative. Plusieurs ont été signalés par les traducteurs qui ont répondu à la consultation.

---

<sup>7</sup> Cette section est rédigée à partir des commentaires laissés par les traducteurs et ne saurait en aucun cas être considérée comme un avis de la Délégation concernant la qualité des outils mentionnés.

Parmi ces outils mentionnés par les répondants, la plupart sont compatibles avec une large diversité d'environnements de travail et de systèmes d'exploitation. Les fonctions non-couvertes peuvent l'être par d'autres logiciels également libres (alignement, formats supplémentaires, pré et post-traitements, édition de mémoires). Les répondants signalent également des outils de terminologie puissants et des fonctionnalités avancées chez certains outils, ainsi que des fonctions d'alignement très efficaces. L'ergonomie reste cependant inégale : appréciable chez certains outils, elle est peu intuitive pour d'autres.

Les traducteurs ont également mentionné plusieurs outils en ligne :

- Wordbee
- XTM
- Boltran
- Crowdin
- Lingotek
- Systran Online Translation pour site internet



On peut distinguer trois grandes familles d'outils selon les hypothèses sur lesquelles ils sont construits. L'une, l'approche *par règles*, suppose que l'on formalise les règles et les dictionnaires des langues à traduire, et que l'on code les algorithmes correspondants dans des programmes informatiques. L'*approche statistique* simule le comportement humain d'apprentissage à partir d'exemples. Enfin, des travaux plus récents adoptent une approche dite *hybride* et combinent, de différentes façons, les approches par règles et statistiques pour bénéficier des avantages des deux, ou du moins pour pallier les inconvénients de chacune.

#### 4.1.1 La traduction par règles

Les premières expériences dans le domaine des approches par règles consistaient à tenter de reconnaître l'organisation grammaticale d'un texte, puis à essayer avec l'utilisation de dictionnaires et de règles de correspondances, de produire un texte traduit<sup>8</sup>.

Exemple : Les traducteurs sont entraînés pour traduire des documents variés.

L'**analyse** de la phrase ci-dessus donne différentes indications : que **Les** et **traducteurs** forment un *groupe nominal*, tout comme **des documents variés**, que ce sont les **traducteurs** qui sont **entraînés**, et non pas **les documents**, que le groupe verbal **sont entraînés** est à la forme passive, etc.

Toutes ces informations sont représentées sous une forme exploitable par un programme informatique. La traduction consiste alors à **transformer** cette représentation pour produire une représentation équivalente dans la nouvelle langue. Cela nécessite de décrire les équivalences entre les mots, et donc de disposer d'un dictionnaire multilingue adapté, par exemple :

**traduire (verbe)** a pour équivalent allemand *übersetzen*, verbe transitif ou intransitif, *über* ne peut pas être séparé de la racine *setzen*.

**Document (nom, masculin)** a pour équivalent allemand *Dokument*, nom, neutre, pluriel *Dokumente*.

---

<sup>8</sup> Le lecteur pardonnera les approximations, il s'agit ici d'une vue d'ensemble.

Il faut également décrire la façon d'arranger ces mots en fonction des propriétés grammaticales de la phrase à produire. Par exemple, traduire de façon formelle, et donc programmable, une règle comme :

**En allemand, les propositions subordonnées demandent que le verbe soit placé à la fin de la proposition.**

Le passage d'une langue à l'autre peut se faire directement : on code les équivalences grammaticales de la langue à traduire vers la langue traduite, ou indirectement en passant par une représentation intermédiaire qui se veut universelle. Cette seconde approche a pour avantage de diminuer l'effort nécessaire pour créer un outil de traduction pour un nouveau couple de langues. La difficulté, cependant, est de parvenir à définir une représentation universelle qui regroupe l'ensemble des propriétés des langues que l'on souhaite traduire.

La dernière étape consiste alors à combiner ces informations pour **générer** un texte dans la nouvelle langue qui soit le plus proche possible d'une production naturelle (produite par un rédacteur humain).

La traduction vers l'allemand de l'exemple cité plus haut, à travers ces trois étapes – analyse, transformation, génération – peut ainsi produire le texte suivant :

**Übersetzer sind trainiert, um verschiedene Texte zu übersetzen.**

(littéralement : traducteurs sont entraînés, pour variés textes à traduire).

Les différents systèmes de traduction automatique à base de règles varient essentiellement par la façon dont ils traitent ces trois étapes : utilisation plus ou moins importante d'un dictionnaire d'équivalents de termes, utilisation de transformations interlingues spécifiques, passage par une représentation intermédiaire indépendante de la langue, richesse des entrées du lexique.

Bien sûr, ces technologies se heurtent avec plus ou moins de succès aux nombreuses difficultés que présentent la langue et la traduction : ambiguïtés des lexiques – **entraîner** est pris ici au sens de *acquérir une compétence par répétition* et non *pratiquer une activité physique régulièrement* –, ambiguïté des phrases – par exemple, dans la phrase *Le traducteur a tancé le relecteur, il était ivre* – qui est ivre ?, différences de structures entre les langues, effets de style...

Ces difficultés sont plus ou moins bien résolues selon les approches adoptées par les systèmes, la connaissance que l'on a des langues traduites et à traduire, la distance entre les langues, les efforts faits pour coder les règles et les lexiques dans les programmes...

### 4.1.2 Les approches statistiques

L'apparition d'internet – et avec elle la disponibilité de grands volumes de données et la nécessité de pouvoir traiter ces données grâce à des ordinateurs de plus en plus puissants – a permis l'éclosion des approches dites statistiques. L'hypothèse qui sous-tend de tels systèmes étant que l'ensemble des connaissances nécessaires pour traduire un texte peuvent être extraites automatiquement à partir d'exemples. Il s'agit cette fois d'entraîner ces systèmes à apprendre à partir d'exemples issus de corpus volumineux déjà traduits puis à produire un texte correspondant le plus possible au texte à traduire.

La première tâche consiste donc à aligner les corpus d'exemple, c'est-à-dire à reconnaître quel extrait de texte (mots ou groupes de mots) d'une langue correspond à tel extrait dans une autre langue. Cet alignement se fait de façon semblable à celle utilisée pour aligner les mémoires de traduction. À partir de cette masse d'exemples bilingues, l'apprentissage consiste pour l'essentiel à calculer la probabilité qu'un segment soit la traduction de tel autre segment dans la langue cible. Produire une traduction consiste alors à proposer la suite la plus probable d'éléments dans une langue cible à partir d'un texte source.

Les systèmes de traduction diffèrent par leur capacité et leur pertinence à mettre les éléments en correspondance, ainsi que par les algorithmes de calcul de probabilités utilisés pour proposer une traduction.

### 4.1.3 Les approches hybrides

Les expériences les plus récentes en traduction automatique – approches dites *hybrides* – combinent approches par règles et approches statistiques. Par exemple, un premier traitement linguistique effectue une analyse pour regrouper les éléments de texte – mots, syntagmes... – sur lesquels porteront les statistiques. Des méthodes statistiques peuvent

également être mises en œuvre pour enrichir des dictionnaires ou des grammaires bilingues utilisées par un système à base de règles. La traduction produite par une approche statistique peut également être redressée a posteriori par une analyse linguistique.

#### 4.1.4 L'évaluation en traduction automatique

Le développement d'outils de traduction automatique étant une tâche ardue, il est nécessaire de vérifier que tout cela fonctionne, et donc d'évaluer la qualité des traductions produites.

L'évaluation doit donc répondre à la question « Quelle est la qualité des traductions produites par l'outil ? ». La réponse n'est cependant pas immédiate. Elle nécessite d'abord de décider ce qu'est une bonne traduction d'un texte. En effet, il existe potentiellement autant de bonnes traductions d'un texte que de traducteurs humains. Comment décider dès lors qu'une traduction produite par un système est de *qualité humaine*, c'est-à-dire qu'elle aurait pu être produite par un rédacteur en chair et en os ? Il faut ensuite mesurer et quantifier l'écart entre une mauvaise et une bonne traduction : telle traduction est-elle plus ou moins bonne que telle autre...

La mesure doit également être répétable, il ne s'agit en effet pas de comparer différentes traductions d'un même document, mais la capacité d'un logiciel à produire de bonnes traductions. Enfin, puisque l'on a besoin d'effectuer de nombreuses mesures, il est préférable que ces évaluations puissent être effectuées à coûts minimaux.

La conception d'une mesure doit également tenter de diminuer les biais, c'est-à-dire l'influence de facteurs externes.

## 4.2 Les avis des traducteurs

**125 traducteurs** ont répondu aux questions de la consultation relatives à la famille d'outils « traduction automatique ». Il s'agit du nombre de répondants le plus faible recueilli pour les catégories proposées lors de la consultation après la « reconnaissance optique de caractères », ce qui indique le faible intérêt des traducteurs professionnels pour ce type d'outils.

Un peu plus d'un tiers des traducteurs (**35%**) ayant participé à la consultation déclarent utiliser au moins un logiciel de traduction automatique et, parmi eux, 10% en utilisent au moins deux. Pour la grande majorité (65%), il s'agit du service en ligne proposé par Google, puis l'outil Systran (10%) et Reverso (7%). Environ 40% utilisent d'autres solutions (MyMemory, Moses ou outil propriétaire).

Le nombre de réponses collectées ne fournit des indications que pour les deux outils les plus utilisés.

## 4.2.1 Ergonomie de la solution

Tableau 17: Évaluez les éléments suivants en choisissant une note de 1 à 5 (5 étant la meilleure note):

	Systran	Google Traduction
L'intégration de la solution de traduction automatique dans votre environnement de traduction	1,5	<b>2,1</b>
L'ergonomie de l'interface de traduction automatique	1,5	<b>2,1</b>
La qualité du support	1,5	1,5
Moyenne	1,5	<b>1.9</b>

## 4.2.2 Spécialisation

Différents moyens peuvent être mis à la disposition du traducteur pour améliorer la qualité de la traduction effectuée, soit en intervenant sur les règles de traduction ou sur la terminologie.

Tableau 18: Pour la solution de TA que vous utilisez, quelle est votre opinion concernant les fonctionnalités suivantes?

	Utilité	Systran	Google Traduction
Possibilité de personnaliser le moteur ou les règles de traduction	4.4	<b>3.3</b>	3
Possibilité de modifier les dictionnaires de traduction	3.8	<b>5</b>	3.8

Prise en compte de la terminologie client	4.5	<b>5</b>	4.4
Possibilité de personnaliser le moteur ou les règles de traduction	4.4	<b>5</b>	4.4
Moyenne	4.3	<b>4.6</b>	3.9

### 4.2.3 API et évolution de la traduction

Une API (*Application Programming Interface*), ou interface de programmation, est un ensemble de fonctions servant de façade par laquelle un logiciel offre un service à d'autres logiciels<sup>9</sup>. Une API est effectivement proposée dans la version payante de Google Traduction et dans l'offre *Translation Server* de Systran.

Tableau 19 : API et évolution de la traduction

	Systran	Google Traduction
La solution propose-t-elle une API?	2.5	<b>3</b>
La qualité de la traduction est-elle améliorée en fonction des textes déjà traduits?	<b>5</b>	2.5
Moyenne	<b>3.8</b>	2.8

## 4.3 Astuces et commentaires de traducteurs<sup>10</sup>

Bien que les traducteurs, dans l'ensemble, estiment que la traduction automatique propose des traductions de *mauvaise qualité*, que ce sont des *solutions générales, sans spécificité sur la terminologie et sans compréhension fine du contexte, donc à utiliser avec beaucoup de parcimonie*, ils trouvent parfois certains avantages à ces outils pour des besoins ponctuels dans le processus de traduction :

<sup>9</sup> Voir [http://fr.wikipedia.org/wiki/Interface\\_de\\_programmation](http://fr.wikipedia.org/wiki/Interface_de_programmation)

<sup>10</sup> Cette section est rédigée à partir des commentaires laissés par les traducteurs et ne saurait en aucun cas être considérée comme un avis de la Délégation concernant la qualité des outils mentionnés.

- *En fait, j'utilise le TA surtout pour vérifier les textes que j'écris : j'écris dans une langue X puis je fais traduire en langue Y et ceci visualise les fautes de frappe de l'original en les transformant en segments incompréhensibles en langue Y. C'est bien plus efficace que le vérificateur d'orthographe - lequel laisse à désirer dans certaines langues.*
- *Malgré ses manquements abondants et évidents, Google Traduction donne instantanément un texte que l'on peut remodeler à sa guise. Pour moi, c'est entrer du texte qui prend le plus de temps.*

### 4.3.1 Les autres produits

Les traducteurs ont également mentionné les produits suivants. On notera que la différence entre traduction automatique et mémoire assistée semble ténue pour les utilisateurs :

- **ETAP-3** : système propriétaire de traitement linguistique spécialisés pour l'anglais et le russe.
- **Libellex** : plateforme de travail multilingue pour l'entreprise.
- **Lingua Custodia** : mémoire de traduction pour la traduction financière
- **MyMemory** : se présente comme *la plus grande mémoire de traduction au monde*.
- **Moses** : logiciel libre de traduction automatique à base de statistiques.
- **Lingway**



équivalents étrangers. C'est aussi le cas au Québec, qui propose, avec le Grand Dictionnaire Terminologique<sup>13</sup>, d'accéder à un fonds terminologique de 3 millions de termes en français.

### 5.1.1 La fiche terminologique

La fiche terminologique est en quelque sorte la *carte d'identité* d'un terme. Elle contient la définition du terme, son équivalent dans une ou plusieurs langues, éventuellement des informations d'usage (le terme est validé, en cours de traitement ou obsolète), des termes proches (synonymes, termes plus génériques ou plus spécifiques, par exemple), ou des exemples d'usage. La fiche terminologique conserve également l'histoire du terme : dans quelles phrases ou quels documents il a été trouvé, qui l'a recueilli, qui l'a validé. Ces informations sont importantes, car elles donnent des indices sur la fiabilité d'une entrée terminologique.

Le travail terminologique consiste pour une grande part à établir un consensus entre terminologues et experts du domaine. Les outils de gestion terminologique partagée facilitent l'élaboration en commun d'une base terminologique et offrent également des fonctions collaboratives telles que les forums, les discussions en ligne ou la gestion de flux de travaux. Les fonctions d'export dans des standards dédiés permettent de diffuser une base terminologique, afin d'homogénéiser les pratiques linguistiques et de mutualiser les efforts entre différents services de traductions.

### 5.1.2 Maintenir une base terminologique

Une base terminologique est un instrument de partage et de capitalisation des connaissances qui doit sans cesse être adapté et enrichi selon les pratiques linguistiques et l'évolution du vocabulaire du domaine. Ce travail de mise à jour est primordial, car il a des conséquences directes sur la qualité et les coûts de la traduction automatique ou humaine. Les logiciels de gestion terminologique proposent pour cela des fonctions de gestion de versions des entrées terminologiques; de même, les fonctions d'extraction terminologique automatique peuvent prendre en compte la terminologie existante.

---

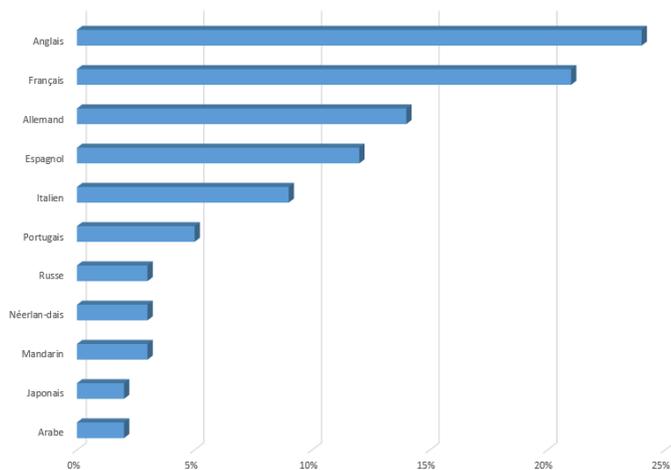
<sup>13</sup> Site internet du Grand Dictionnaire Terminologique : <http://www.granddictionnaire.com/>

## 5.2 Les avis des traducteurs

**267 traducteurs** ont répondu aux questions de la consultation relatives à la famille d'outil de « gestion terminologique », parmi lesquels **48 %** ont déclaré utiliser au moins un outil de ce type, ce qui signifie que leur usage semble relativement courant. C'est la catégorie d'outils proposée lors de la consultation qui a recueilli le plus grand nombre de réponses.

Bien que la consultation ait été conduite pour des traducteurs francophones, l'anglais est la langue pour laquelle le travail de terminologie est le plus fréquent. Notons que, selon le baromètre du poids des langues<sup>14</sup>, les langues les plus traduites dans le monde sont, dans l'ordre, l'*anglais*, le *français* puis l'*allemand*, et les langues vers lesquelles les textes sont traduits sont, dans l'ordre, l'*allemand*, l'*espagnol* puis l'*anglais*.

Illustration 6 : Quelles sont les langues pour lesquelles votre service est habituellement amené à gérer de la terminologie (plusieurs réponses possibles) ?



### 5.2.1 Les logiciels

Trois outils de gestion terminologique ont été mentionnés par suffisamment de traducteurs pour que les résultats soient exploitables :

<sup>14</sup> <http://wikilf.culture.fr/barometre2012/>, consulté le 5 janvier 2014

Produit	SDL Multiterm	MemoQ	TermStar
Éditeur	SDL	Kilgray Translation Tech	STAR Group
Pays	Royaume-uni	Hongrie	Suisse
Site	www.sdl.com	www.memoq.com	www.star-group.net

## 5.2.2 Les fonctionnalités

Les outils de gestion terminologique offrent un large spectre de fonctions, de la saisie et la personnalisation de fiches terminologiques à la collaboration avec des experts.

Tableau 20 : Selon votre expérience et vos besoins, quel est votre avis concernant les fonctionnalités suivantes ?

	Utilité	SDL Multiterm	MemoQ	TermStar
Vérification de doublons potentiels	3.4	4.2	5	3.3
Possibilité de poser un indicateur de fiabilité sur les entrées terminologiques	2.9	4.5	5	3.8
Possibilité d'indiquer des termes proscrits	3.7	4.6	5	5
Filtre de la base terminologique par date	3.4	5	5	3.8
Possibilité de paramétrer la fiche terminologique	4.2	4.6	5	5
Possibilité de lier les fiches terminologiques entre elles	4	3.5	5	5
Possibilité d'associer des documents multimédia à une fiche terminologique	2.5	4.2	5	5
Possibilité de partager tout ou partie de la terminologie avec des partenaires ou des sous-traitants	3.3	4.2	5	5
Possibilité de poser des commentaires sur une fiche terminologique afin d'obtenir un consensus avec des terminologues ou des experts	3.4	5	5	3.8
Possibilité de faire valider une fiche terminologique par un expert du domaine	3.9	5	5	5
Possibilité d'importer des glossaires existants	4.7	4.6	5	5
Accès à des ressources externes spécialisées (dictionnaires, bases terminologiques grand public, etc.)	3.5	4	5	5
Le traducteur peut mettre à jour la terminologie pendant qu'il traduit un document	4.9	4.4	5	5
Moyenne	3.6	4.5	5	4.6

### 5.2.3 Formats d'import et d'export

Une bonne gestion terminologique garantit la cohérence du vocabulaire des textes traduits, mais également des textes rédigés au sein d'un organisme. Elle est donc essentielle, tant pour la clarté des textes que pour l'image de l'entreprise ou de l'administration. Il est donc nécessaire de pouvoir la diffuser auprès des traducteurs, des collaborateurs et des différents services. Différents formats d'échange sont possibles, avec chacun leurs spécificités :

- CSV (pour *Coma Separated Values*) : format texte très facilement transportable d'un logiciel à l'autre, y compris des logiciels bureautiques tels que les tableurs.
- XML (pour *eXtended Mark Up Language*) : format facilement transportable d'un logiciel à l'autre, potentiellement très riche en informations et très structuré.
- TBX (pour *Term Base eXchange*) : format standardisé (ISO 30042:2008) pour la représentation et l'échange de terminologies conceptuelles.

Tableau 21 : Quels sont les formats d'importation et d'exportation proposés par l'outil de gestion terminologique ?

Import	SDL Multiterm	MemoQ	TermStar
CSV	1.9	4.3	<b>5</b>
XML	1.5	<b>2.9</b>	2
TBX	1.7	0.7	<b>2</b>
Aucun	0.1	-	-
XLS	0.4	<b>0.7</b>	-
txt	0.1	-	-
Export	SDL Multiterm	MemoQ	TermStar
CSV	1.7	<b>5</b>	<b>5</b>
XML	0.8	<b>2.9</b>	2
TBX	1.5	-	<b>2</b>

## 5.2.4 Support et ergonomie

Tableau 22 : En ce qui concerne la solution de gestion terminologique, choisissez une note de 1 à 5 (5 étant la meilleure note)

Import	SDL Multiterm	MemoQ	TermStar
l'ergonomie de l'interface fournie au traducteur	2.6	3.4	<b>3.6</b>
le service-support assuré par le fournisseur	1.3	<b>2.9</b>	2.4

## 5.3 Astuces et commentaires de traducteurs<sup>15</sup>

Les traducteurs ont également mentionné les produits suivants : *Babylon Glossary Builder*, déjà vu, et les éditeurs de texte. En ce qui concerne ces derniers, l'utilisation de glossaires au format texte offre une grande souplesse : choix des champs, utilisation des fonctions de programmation (recherche, tri, concaténation de fichiers, élimination des doublons, etc.).

---

<sup>15</sup> Cette section est rédigée à partir des commentaires laissés par les traducteurs et ne saurait en aucun cas être considérée comme un avis de la Délégation concernant la qualité des outils mentionnés.



## 6. Extraction terminologique

### 6.1 Comment ça marche?

L'extraction terminologique consiste, à partir de textes disponibles sous forme électronique dans différents *formats*, après un éventuel *prétraitement*, à identifier des candidats-termes. Ces outils sont essentiellement basés sur deux *principes de fonctionnement* – extraction par règles ou par chaîne de caractères répétées. Leur qualité peut être appréciée selon la qualité de l'extraction, globalement ou par langue, ainsi que par les *fonctions* qu'ils proposent. Une fois l'extraction effectuée, les candidats-termes peuvent être présentés sous la forme d'un concordancier ou d'un réseau de termes, puis exportés dans différents formats.

#### 6.1.1 Les outils d'extraction terminologique

Le terminologie rassemble un glossaire des termes spécifiques par un examen minutieux des textes d'un domaine. Il peut être assisté dans cette tâche par un outil d'extraction terminologique automatique, qui repose sur des méthodes linguistiques ou statistiques. Les méthodes linguistiques repèrent des suites de mots répondant à certains critères grammaticaux, par exemple, un *nom* suivi d'un *participe passé* qui reconnaîtra ainsi, par exemple, dans le texte ci-dessous, les termes (ou, plus exactement, les *candidats-termes*) soulignés :

### Exemple :

Nous avons eu recours à un logiciel de [traduction assistée](#) pour ce document. La [traduction corrigée](#) par un [traducteur spécialisé](#) a ensuite été revue par deux [experts désignés](#) du domaine. Puis cette [traduction révisée](#) a ensuite été mise en forme pour une [diffusion étendue](#).

Évidemment, les critères grammaticaux doivent être plus précis, comme exiger que la suite de mots soit précédée par un article, ce qui, dans l'exemple précédent, écarte trois des candidats-termes. La multiplication de ces critères aboutit à une description linguistique précise des termes extraits par le logiciel.

La seconde approche repose sur l'hypothèse que le vocabulaire de spécialité est plus fréquent dans les textes du domaine. Il semble par exemple logique que le terme *réacteur nucléaire* soit plus probable dans des textes traitant des énergies que dans un autre qui traiterait de la littérature française du 19<sup>e</sup> siècle. Les outils d'extraction automatique statistique combinent donc différents indicateurs statistiques pour identifier les candidats-termes spécifiques à un domaine et les proposer au terminologue qui décidera éventuellement de les ajouter à la terminologie du domaine. Ces outils d'extraction peuvent également identifier les acronymes présents dans les textes, et reconnaître automatiquement leur forme développée.

L'outil d'extraction terminologique automatique affiche ces termes dans un concordancier, monolingue ou multilingue, avec le contexte dans lequel chacun des termes proposés a été trouvé. Cette information donne de précieux indices au terminologue et au traducteur sur le sens de ces termes en contexte.

## 6.1.2 Les outils d'extraction terminologique multilingue

Les outils d'extraction terminologiques sont également conçus pour analyser des textes déjà traduits, voire des textes d'un même domaine dans des langues différentes, et pour proposer les termes avec leur équivalent dans une autre langue. Ces outils adoptent pour cela différentes stratégies, dont l'alignement des textes ou le repérage d'équivalences présentées plus haut. Ces fonctions d'analyse présentent l'avantage d'afficher les termes et leurs équivalents dans des concordanciers multilingues qui apportent un gain de temps important dans la constitution d'une terminologie multilingue.

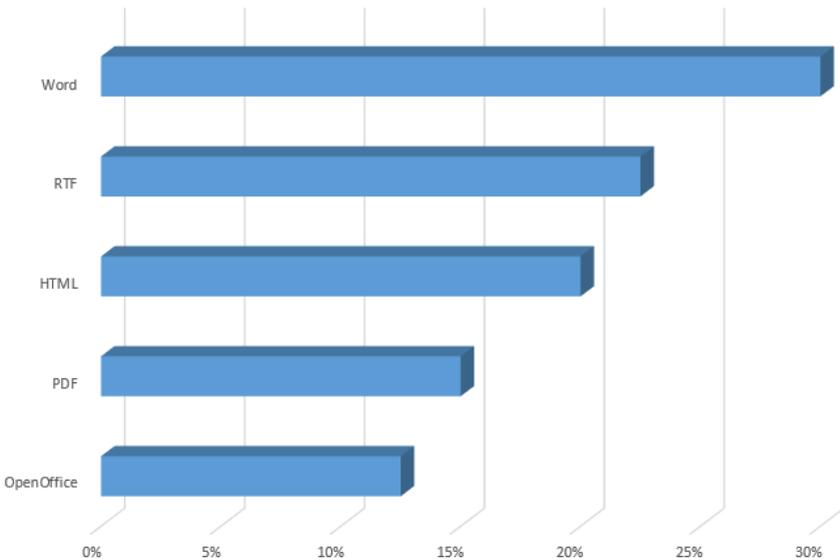
## 6.2 Les avis des traducteurs

**196 traducteurs** ont répondu aux questions de la consultation relatives à la famille d'outils d'« extraction terminologique » mais seulement **17%** déclarent les utiliser. Les outils d'extraction terminologique automatique, bien qu'ils soient proposés par les fournisseurs, semblent donc peu utilisés par les traducteurs.

*SDL MultiTerm Extract*, outil utilisé le plus souvent, ne recueille que quatre réponses. Les données recueillies pour cet outil ne sont donc pas suffisamment représentatives pour être reportées dans ce document.

### 6.2.1 Formats des fichiers analysés

Illustration 7 : La solution extrait des termes dans des documents au format suivant :



L'extraction de candidats-termes est faite à partir de documents dans différents formats. La suite Microsoft Office arrive là encore largement en tête des formats traités par les différents outils selon les traducteurs interrogés. LibreOffice/OpenOffice, sous licence libre, n'arrive qu'après le format PDF.

## 6.2.2 Prétraitement

Les outils d'extraction ne nécessitent un prétraitement que pour un quart des répondants et sont donc dans la majorité des cas à même d'opérer sur des fichiers bruts.

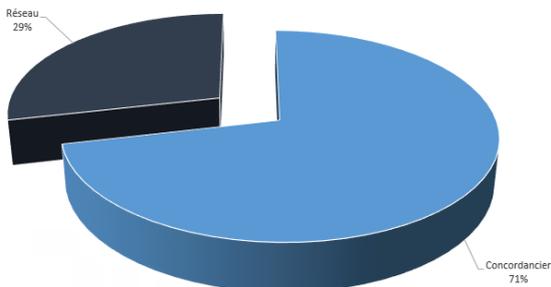
Illustration 8 : Les textes doivent-ils être traités préalablement à l'extraction terminologique ?



## 6.2.3 Principe de fonctionnement

Les solutions d'extraction terminologique se répartissent en deux grandes catégories, selon qu'elles fonctionnent par reconnaissance de motifs d'extraction ou par identification de chaînes de caractères répétées. Le mode de fonctionnement de ces outils n'est cependant que rarement accessible au traducteur.

Illustration 9 : Sur quels principes la solution fonctionne-t-elle ?



## 6.2.4 Qualité de l'extraction et du service associé

Si les fonctions d'extraction semblent être appréciées des traducteurs, c'est également le cas de la qualité d'extraction des termes et du service fourni avec l'outil.

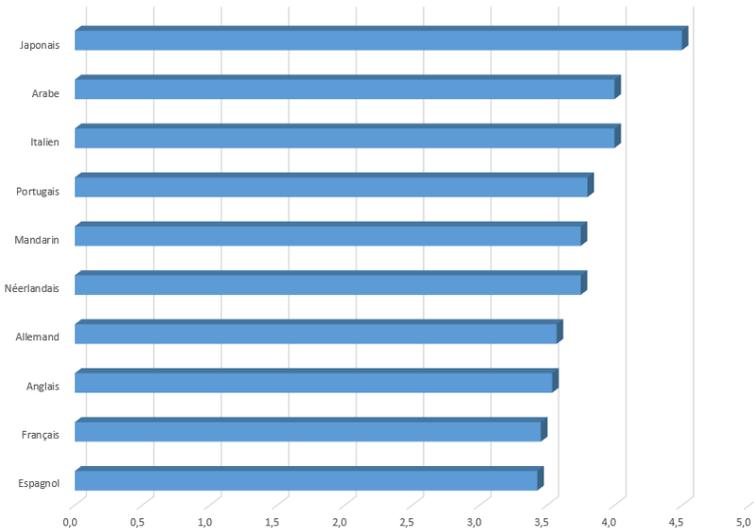
Tableau 23 : Évaluez les éléments suivants en choisissant une note de 1 à 5 (5 étant la meilleure note)

La qualité des termes proposés après une extraction terminologique	3,2
L'ergonomie de l'interface de l'outil d'extraction terminologique	3,3
La qualité du service support du fournisseur de la solution	3,8

## 6.2.5 Qualité par langue

Les solutions d'extraction terminologique sont perçues comme étant applicables à tous types de langues.

Illustration 10 : Évaluez la qualité de l'extraction terminologique en choisissant de 1 à 5 (5 étant la meilleure note)



## 6.2.6 Fonctions d'extraction

Les données ci-dessous concernent, sans les séparer, l'ensemble des outils d'extraction terminologique mentionnés par les traducteurs qui ont répondu à l'enquête. Considérant le faible échantillon de réponses, ces données ont été rassemblées et sont à prendre avec une grande précaution.

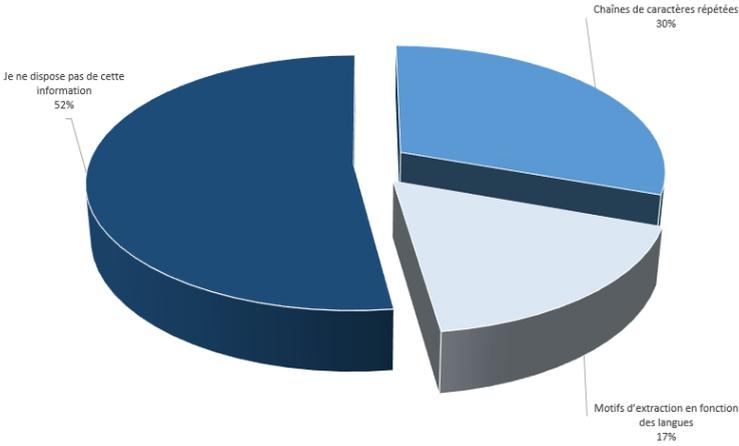
Tableau 24 : En fonction de votre expérience, quel est votre avis concernant les fonctionnalités suivantes ?

	Utilité	Score
Le terminologie, administrateur ou utilisateur de la solution, peut modifier les règles d'extraction	4	4.4
La solution prend en compte une terminologie existante lors de l'extraction des candidats-termes	4	5
La solution d'extraction extrait les sigles et leurs extensions	5	5
La solution extrait des termes composés	4.6	5
Les termes extraits sont normalisés (seules les formes canoniques sont présentées : pluriel réduit au singulier, notamment)	3.3	5
La solution extrait des définitions ou des collocations	4	5
La solution extrait des candidats-termes bilingues	4.2	4.2
La solution recherche des traductions possibles des candidats-termes dans une mémoire de traduction	5	4.4

## 6.2.7 Restitution des résultats d'extraction

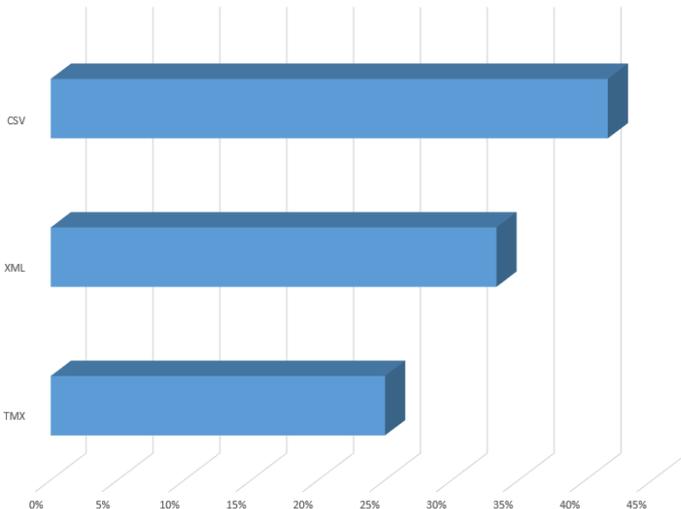
Les résultats d'extraction sont, dans la plupart des cas, restitués sous la forme d'un concordancier qui affiche les termes extraits avec les contextes dans lesquels ils ont été trouvés. Les termes peuvent être également affichés sous la forme d'un réseau explicitant les relations entre les termes.

Illustration 11 : Sous quelle forme les candidats-termes sont-ils affichés ?



### 6.2.8 Exports des candidats-termes

Comme il a été déjà mentionné, le partage de la terminologie favorise l'homogénéisation des vocabulaires utilisés dans les documents traduits et rédigés au sein d'un organisme, et est un facteur de gain de temps. La question de l'export des candidats-termes vers d'autres outils est donc essentielle. Seul un quart des solutions permet l'export au format standard de gestion de terminologie.





## **Ministère de la Culture et de la Communication**

### **Délégation générale à la langue française et aux langues de France**

6 rue des Pyramides

75001 Paris

téléphone: 01 40 15 73 00

télécopie: 01 40 15 36 76

courriel: [dglflf@culture.gouv.fr](mailto:dglflf@culture.gouv.fr)

[www.culturecommunication.gouv.fr/Politiques-ministerielles](http://www.culturecommunication.gouv.fr/Politiques-ministerielles)

[/Langue-francaise-et-langues-de-France](http://Langue-francaise-et-langues-de-France)

[twitter.com/languesFR](https://twitter.com/languesFR)

### **Délégué général**

Loïc Depecker

### **Délégué général adjoint**

Jean-François Baldi

### **Coordination générale du projet**

Thibault Grouas, chef de la mission des langues et du numérique, DGLFLF

### **Conduite de la consultation publique et réalisation du guide**

Alain Couillault, chargé de mission innovation, Aproged & Professeur associé, université de La Rochelle

Gaïd Evenou, chargée de mission pour la traduction, DGLFLF

Lucie Gianola, chargée de projets, DGLFLF

Quentin Samier, chargé de projets, DGLFLF

### **Coordination éditoriale**

Pauline Chevallier

### **Graphisme**

Claire Méry



*Liberté • Égalité • Fraternité*

RÉPUBLIQUE FRANÇAISE





*Liberté • Égalité • Fraternité*  
RÉPUBLIQUE FRANÇAISE



## Délégation générale à la langue française et aux langues de France

6 rue des Pyramides, 75001 Paris

téléphone : 01 40 15 73 00

télécopie : 01 40 15 36 76

courriel : [dglff@culture.gouv.fr](mailto:dglff@culture.gouv.fr)

[www.culturecommunication.gouv.fr/Politiques-](http://www.culturecommunication.gouv.fr/Politiques-)

[ministerielles/Langue-francaise-et-langues-de-France](http://www.culturecommunication.gouv.fr/Politiques-)

[twitter.com/languesFR](https://twitter.com/languesFR)