

Peut-on traduire l'identité? De la différence entre textes traduits et textes originaux en Suisse romande

Alice DELORME BENITES

Université des Sciences appliquées de Zurich – ZHAW

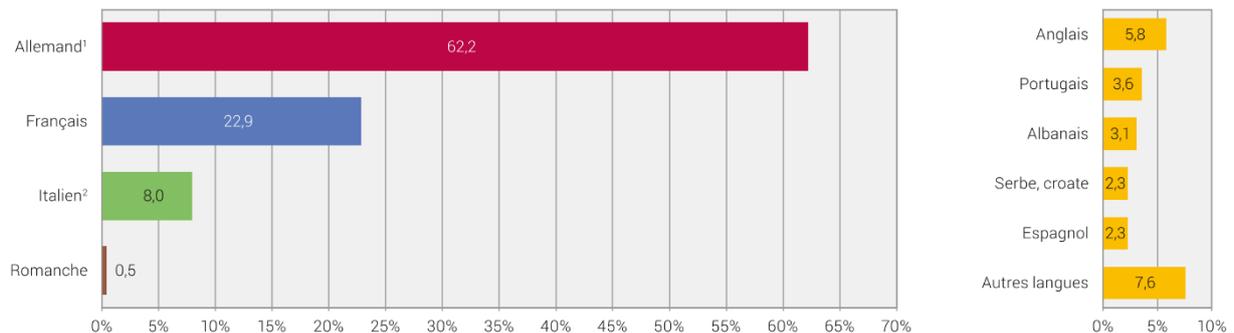
As a multilingual country, Switzerland heavily relies on translation for countrywide public communication, for companies as well as for governmental institutions. All, but especially the latter, face the challenge of providing all citizens with the same information, regardless of the language they speak. This study investigates a possible influence of German, the main language in major companies and institutions in Switzerland, on Swiss-French at the discourse level. This influence, resulting from a disproportional number of French texts being translations from German, is evidenced by major differences concerning the topics they address. Evidence is found by investigating a large text corpus, Swiss-AL: keywords and toponyms contained in translated vs. original texts show differences in the weight of the various levels of political administration and in the associations made with specific places in Switzerland. These differences go beyond the expected contrast between nation-wide and local texts. This research has been conducted with texts mostly written before the explosion of neuronal machine translation, which arises the question of whether the newest improvement and instant availability of automatic translation might increase exponentially this influence.

1. Introduction: allemand, français et traduction en Suisse

"Pour les Romands, c'est la même chose mais en français" (Donzé 2020). Cette indication fait office d'instruction d'usage sur un tube de crème hydratante commercialisée en Suisse. S'il s'agit (fort heureusement), d'une boutade de deux jeunes entrepreneurs romands essentiellement destinée à amuser le public francophone, elle met en évidence l'image que se font les Romands des mécanismes de communication publique à l'échelle nationale. Non seulement les minorités linguistiques (italophones, francophones, romanchophones) lisent au quotidien de nombreuses traductions de l'allemand, langue maternelle de quelque 62% de la population (voir Figure 1), mais le concept-même de traduction semble pris au sens étroit du terme, sans adaptation culturelle particulière: "c'est la même chose". Les textes traduits sont de diverses natures, soit qu'ils aient dès le départ une visée nationale (textes fédéraux, par exemple), soit qu'ils soient à l'origine destinés à un public germanophone puis élargis aux autres régions linguistiques (textes d'entreprises et d'universités, par exemple). Cette surreprésentation de l'allemand est souvent perçue comme un problème, notamment au sein des institutions fédérales, où l'on déplore encore des effectifs trop faibles de fonctionnaires francophones et italophones (Kübler et al. 2020; d'Urso 2014). Bien qu'officiellement, l'allemand, le français et l'italien figurent sur un pied d'égalité au sein de l'administration fédérale, l'allemand prédomine. Ainsi, une enquête sur les processus de rédaction des textes au sein des instances fédérales et des institutions des cantons bilingues montre que les textes sont le plus souvent produits d'abord en allemand, puis traduits

vers le français et l'italien (Stoffel 2019), et de nombreuses voix s'élèvent pour dénoncer la prépondérance d'un "esprit" allemand dans les textes juridiques dès leur ébauche (Arquint 2014: 137). Ces observations rejoignent, en quelque sorte, l'ironie des producteurs de crème romands: la domination de l'allemand comme langue de création originale des textes et discours publics pose un problème, engendre une frustration. Mais ne s'agit-il pas là d'une situation conditionnée par le multilinguisme?

Langues déclarées comme langues principales, en 2018



¹ ou suisse allemand

² ou dialecte tessinois, italo-grison

Population résidante permanente vivant en ménage privé. Les personnes interrogées pouvaient indiquer plusieurs langues principales.

Source: OFS – Relevé structurel (RS)

© OFS 2020

Figure 1. Répartition des langues en Suisse en 2018, Office fédéral de la statistique 2020

Il convient alors d'esquisser brièvement les contours du multilinguisme suisse. En effet, la Suisse "n'est pas un État national centralisateur classique, ayant renoncé à une langue nationale unique qui serait liée à l'identité nationale." (Manno 2007 §1¹). Dans ce contexte, elle se distingue surtout par un plurilinguisme *institutionnel* (l'allemand, le français, l'italien et le romanche sont des langues nationales et officielles) qui, paradoxalement, pérennise l'existence d'un monolinguisme *individuel* (cf. Manno *ibid.*): en 2014, à peine 2% de la population indiquait être bilingue français-allemand (SRF 2016). En effet, si la Confédération telle qu'elle existe aujourd'hui est multilingue, sur 26 cantons, seuls 3 sont bilingues (Fribourg, le Valais, Berne) et les Grisons constituent l'unique canton trilingue de Suisse. L'allemand, langue dans laquelle s'exprime la majorité de la population Suisse et sont rédigés la majorité des textes du corpus Swiss-AL (Krasselt et al. 2020), n'est donc pas officiellement représentée (et à peine pratiquée) dans toute une partie du pays. Dans ce sens, Busekist (2007: 113), décrit la Suisse comme une simple juxtaposition de zones unilingues plutôt qu'un État multilingue de fait. La communication interlinguistique au sein d'un même pays revêt alors une importante dimension interculturelle. On peut donc se demander dans quelle mesure s'opère un certain décalage thématique entre des textes produits en allemand puis traduits

¹ Il s'agit d'un texte en ligne, sans pages numérotées.

vers le français (ou rédigés en français par des locuteurs allemands) et des textes produits directement en français sans influence de l'allemand.

2. Matériau de l'étude: corpus, logométrie, analyse du discours

Notre questionnement présente un double défi méthodologique: émergeant en partie de données fondées (Stoffel 2019) mais aussi d'impressions subjectives (telle la boutade des producteurs de cosmétiques romands), il requiert une démarche aussi objective que possible pour y répondre. Par ailleurs, l'analyse textuelle de quelques textes pris au hasard ne saurait décrire un phénomène à large échelle. Nous avons donc choisi une analyse logométrique sur un corpus numérique existant, qui permet d'inclure un très grand nombre de textes et de garantir un degré d'objectivité élevé. Mayaffre (2005) distingue l'approche logométrique de la lexicométrie car elle va plus loin qu'un simple décompte des lexèmes au sein d'un corpus: il s'agit aussi de considérer les lemmes et les séquences syntaxiques pour accéder à "toutes les unités essentielles du discours" (Mayaffre 2005: 92). Il met d'ailleurs en lumière le potentiel de la logométrie pour notre questionnement:

Avec la logométrie, l'analyste part toujours de faits linguistiques matériels massivement attestés (y compris dans leur absence ou leur rareté) pour commencer à réfléchir et le cas échéant inférer un positionnement idéologique ou une attitude politique. (Mayaffre 2005: 104)

Ainsi, la logométrie peut-elle être prise comme "une méthode qui permet de décentrer l'analyste de ses habitudes de lecture, et de repousser l'entrée dans la subjectivité que toute interprétation requerra" (Mayaffre 2007: 171).

La présente étude se base donc sur l'interrogation logométrique d'un grand corpus numérique multilingue, (Swiss-AL "Swiss-Angewandte Linguistik", en français *Swiss-Linguistique-Appliquée*, Krasselt et al. 2020), qui regroupe des textes extraits de sites Internet du domaine suisse (.ch)². Il s'agit en réalité de quatre corpus non alignés: allemand, français, italien, anglais³. Une comparaison directe de fréquences entre les différentes langues aurait été peu pertinente ici. En effet, une grande partie des textes français étant des traductions fidèles de l'allemand, les thèmes abordés y sont les mêmes et le vocabulaire ne variera qu'à l'échelle de la synonymie. Enfin, il serait erroné de modéliser un discours "francophone" composé en grande part de textes produits, à l'origine, par des germanophones. Ainsi était-il plus approprié de

² C'est par ailleurs dans le cadre d'un projet de recherche sur les discours énergétiques en Suisse (Stücheli-Herlach, Ehrensberger-Dow & Dreesen 2018) qu'a émergé la présente étude. Ce projet de recherche, mené de 2017 à 2019 et financé par l'OFEN, s'attachait à mettre en avant les différents discours publics sur l'énergie en Suisse à l'heure de nombreux changements et décisions dans ce domaine.

³ Le corpus anglais, de taille très réduite, n'est pas utilisé dans cette étude, ni dans le projet de recherche plus large au sein duquel elle s'inscrit. Nous nous contenterons donc de le mentionner ici.

comparer les textes français rédigés à l'origine dans la langue de Molière, et les textes français produits par traduction d'un texte allemand ou rédigés dans un contexte essentiellement germanophone (similitraduction). Le présent article propose donc une étude comparative de deux sous-corpus de Swiss-AL: l'un constitué de textes traduits ou "similitraduits" depuis l'allemand, l'autre constitué de textes rédigés directement en français. Nous pourrions ainsi identifier de possibles différences, qui expliqueraient peut-être la frustration autour de l'abondance des traductions depuis l'allemand dans la sphère publique. Notre approche logométrique repose sur trois étapes, décrites par Mayaffre (2005: 154): la création de sous-corpus à partir du corpus général Swiss-AL, le choix des unités linguistiques à explorer, puis l'interrogation des sous-corpus sous plusieurs formes.

2.1 *Composition du corpus*

Les sites Internet dont sont tirés les textes du corpus Swiss-AL relèvent de divers domaines: politique, économique, scientifique et médiatique. Le corpus inclut notamment les textes publiés sur le site de la Confédération, sur ceux des principaux partis politiques, des grandes et moyennes entreprises, des universités et hautes écoles, mais aussi de la presse généraliste et spécialisée, rassemblant ainsi 301 sources différentes. Swiss-AL compte parmi les plus grands corpus numériques textuels de Suisse. Il est annoté (lemmatisation, catégories grammaticales, entités nommées) et actualisé plusieurs fois par an. Fin 2018⁴, le corpus Swiss-AL allemand comptait 590 millions de lexèmes pour 1,45 millions de textes. Le corpus français rassemblait 420 millions de lexèmes pour presque un million de textes. Enfin, le corpus italien comptabilisait un peu moins de 200 millions de lexèmes pour 500 000 textes.

2.2 *Élaboration des sous-corpus de travail*

Nous avons créé deux sous-corpus regroupant, d'une part, les textes traduits ou similitraduits depuis l'allemand (textes "influencés"), et d'autre part, les textes rédigés directement en français sans influence notable de l'allemand (textes "originaux"). La répartition des textes dans ces deux sous-corpus ne s'est pas faite sur la base du texte lui-même (résultat) mais sur celle de sa rédaction (processus). Il s'agissait de déterminer dans quelle proportion le processus d'écriture s'est effectué sous l'influence de l'allemand. L'absence de frontière nette entre les deux catégories impliquait des décisions qualitatives lors de la création des sous-corpus. Une analyse détaillée de chaque texte individuel aurait été impossible au vu de la taille des corpus, c'est pourquoi nous avons choisi une approche basée sur la reconstruction des processus (*reverse-engineering*) selon Chestermann (2013): nous avons recherché, sur le site

⁴

C'est sur cette version du corpus que repose cet article.

Internet de chaque source présente dans le corpus, des indices permettant de reconstituer le contexte du processus de rédaction des textes. Pour la grande majorité des sources, l'observation minutieuse du site Internet suffisait à une catégorisation sans équivoque. En cas de doute, nous avons contacté directement la source et, si le doute subsistait, la source n'a été intégrée à aucun des sous-corpus. Nous noterons ici que si Mayaffre (2005) propose une approche logométrique conclue par un retour qualitatif au texte lui-même, nous avons, pour cette étude, quelque peu inversé l'ordre en procédant tout d'abord à une analyse qualitative afin d'élaborer les sous-corpus de travail, puis à une analyse quantitative de ces sous-corpus. La taille des sous-corpus ainsi générés constitue d'ores et déjà un résultat (Figure 2). Pour le français, le sous-corpus "originaux" regroupe des textes de presse, des publications communales et cantonales, et des textes d'entreprises, universités et institutions romandes. Le sous-corpus "influencés" rassemble, outre les textes de l'administration fédérale, des publications cantonales et communales, des textes d'entreprises, universités et institutions alémaniques – à portée nationale ou locale.

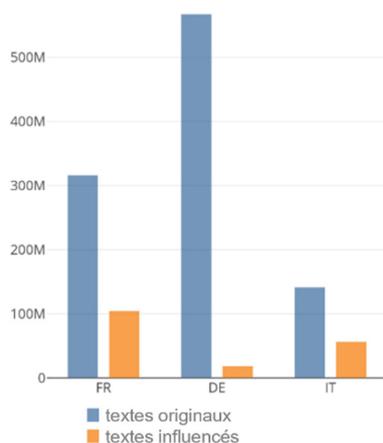


Figure 2. Comparaison de la taille des sous-corpus en millions de tokens

Les textes du corpus allemand ont été très majoritairement catégorisés comme originaux, c'est-à-dire considérés comme rédigés sans influence notable du français ou de l'italien. Ainsi, lors de la rédaction de textes allemands en Suisse, les autres langues nationales semblent exercer une influence mineure; or, la part officielle de germanophones sur le territoire helvétique n'étant que de 62,2% (OFS, 2020), on pourrait y voir une surreprésentation de l'allemand dans les textes publiés sur Internet en Suisse. Se pose alors la question des effets d'une telle disproportion sur les pratiques langagières et le discours⁵. Enfin, l'analyse des sources a fourni un dernier résultat: les textes français influencés

⁵ Nous définirons ici le discours selon les réflexions de Maingueneau (1979) à l'intersection de sa troisième et cinquième acception: "énoncé linguistique intégré à un acte d'énonciation" et "énoncé considéré du point de vue du mécanisme discursif qui le conditionne". (p. 3f.)

ne le sont que par l'allemand. Le corpus ne contient pas de textes rédigés d'abord en italien puis traduits vers le français.

2.3 *Choix des unités de texte et de recherche*

Les interrogations de corpus présentées ici relèvent de trois types: l'analyse des spécificités, qui permet une approche exploratoire guidée par les corpus; l'analyse des toponymes mentionnés dans les corpus, grâce au balisage préalable de ceux-ci, approche également guidée par le corpus⁶; l'analyse complémentaire des co-occurrences de certains résultats des deux précédentes interrogations, permettant un approfondissement basé sur corpus.

Cette étude s'intéresse essentiellement à l'aspect sémantique des unités et se concentre donc sur les mots graphiques. Cette contrainte est en partie imposée par l'outil de travail, CQP web, par exemple pour la recherche de spécificités: la fonction *keywords* correspondante dans cette interface fournit une liste de lexèmes (sans possibilité d'obtenir des *n-grams*). Toutefois, nous avons intégré des enchaînements de mots dans les analyses complémentaires basées sur corpus lorsque c'était nécessaire.

3. **Analyse des spécificités des sous-corpus**

L'analyse des spécificités (en anglais *keywords*), c'est-à-dire "du vocabulaire objectivement spécifique (caractéristique) d'une partie du corpus par rapport à l'ensemble" (Mayaffre, 2007) permet une comparaison des différents sous-corpus générée par les données. Ici, les spécificités sont définies comme des lexèmes dont la fréquence au sein d'un sous-corpus (par exemple "textes influencés") est remarquablement plus élevée que dans l'autre corpus (par exemple "textes originaux"), pris comme référence. Soulignons que, par conséquent, l'absence d'un terme dans ces listes ne signifie pas qu'il soit peu ou pas présent dans le corpus, mais que sa fréquence n'est pas distinctive dans ce corpus par rapport à un autre. L'analyse suivante portera donc sur la comparaison des sous-corpus "influencés" et "originaux", pour le corpus Swiss-AL. Les spécificités sont représentées sous forme de nuages de mots, la taille des caractères variant selon la fréquence relative de chaque lexème.

3.1 *Comparaison générale*

Probablement en raison de la taille et de la diversité thématique du corpus Swiss-AL, les 50 premiers lexèmes ne fournissent pas de tableau homogène (voir Figure 3 et Figure 4). Toutefois, plusieurs champs sémantiques se dessinent

⁶ Le choix d'inclure ici les toponymes repose sur les premières observations des spécificités des sous-corpus.

lexèmes spécifiques aux textes originaux semblent, à première vue, également relever de la politique nationale: *Président* (29), *ministre* (30), *gouvernement* (31). Or, une analyse plus approfondie réfute cette hypothèse et illustre bien la complexité du décalage entre les deux sous-corpus.

3.2 *Président*



Figure 5. Co-occurrences de *Président* dans les textes originaux



Figure 6. Co-occurrences de *Président* dans les textes influencés

La Figure 5 indique une nette tendance des textes originaux à associer le terme à des personnalités étrangères ou liées à l'étranger. Les extraits suivants en attestent:

- (1) Le **président** turc Recep Tayyip Erdogan les a mis en garde s'ils s'en prenaient aux populations turkmènes de Tal Afar. (*textes originaux*)
- (2) Les banques étrangères établies en Suisse s'efforcent de trouver des solutions avec leurs clients afin de régulariser leur situation fiscale, a indiqué Alfredo Gysi, le **président** en partance de l'Association des banques étrangères en Suisse (ABES). (*textes originaux*)

Les textes influencés (Figure 6) affichent eux aussi une certaine propension à lier *président* et étranger, mais moins forte et surtout contrebalancée par la présence de termes d'identification nationale comme *Confédération* ou *Suisse*.

- (3) En sa qualité de "primus inter pares", le **président** de la Confédération assume toutefois par tradition un certain nombre de tâches et de fonctions pendant son année de présidence. (*textes influencés*)

Enfin, on trouvera dans les co-occurrences plus d'indices d'un emploi du terme en lien avec des conseils d'administration (donc dans une dimension plus économique que politique).

3.3 *Gouvernement*

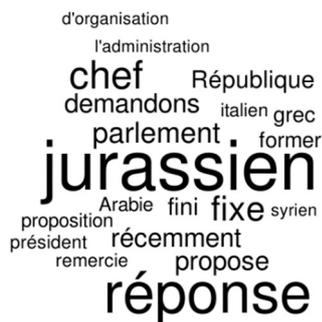


Figure 7. Co-occurrences de *Gouvernement* dans les textes originaux



Figure 8. Co-occurrences de *Gouvernement* dans les textes influencés

Dans les deux sous-corpus, le lexème *Gouvernement* est souvent employé en association avec un pays étranger (Figure 7 et Figure 8).

- (4) Samedi, lors d'un meeting de campagne, l'ancien chef du **gouvernement** italien est revenu sur ses attaques de 2003 [...] (textes originaux)

Le lexème *République*, très présent dans les co-occurrences des textes influencés, est généralement associé à l'étranger également, comme le montre l'exemple ci-après:

- (5) En signant la présente convention, le gouvernement de la Confédération suisse et le **gouvernement** de la République fédérale d'Allemagne tombent d'accord sur les points suivants. (textes influencés)

La différence principale réside donc dans la prédominance de *jurassien* dans les textes originaux, qui fait écho à la dimension locale esquissée lors de l'analyse générale des spécificités. D'ailleurs, la présence de *Confédération* parmi les co-occurrences de *Gouvernement* dans les textes influencés souligne, elle aussi, la dimension plus nationale déjà mise en lumière lors de l'analyse générale.

- (6) Le Jura bernois veut des garanties du **Gouvernement** jurassien avant de se prononcer sur un vote concernant la Question jurassienne. (textes originaux)

3.4 *Ministre*

Les co-occurrences du lexème *Ministre* rejoignent les observations précédentes: la part belle est faite à l'étranger (notamment avec *Premier*, puisqu'il n'existe pas de Premier ministre en Suisse), surtout dans les textes originaux, qui mentionnent une plus grande diversité de pays (*britannique*, *grec*, *israélien*, *français*).

- (1) "Je me demande si elle est durable", avait-il dit lors d'une conférence de presse commune à Londres avec le Premier **ministre** britannique David Cameron. (textes originaux)



Figure 9. Co-occurrences de Ministre dans les textes originaux



Figure 10. Co-occurrences de Ministre dans les textes influencés

Les textes influencés, cependant, mentionnent également volontiers des personnalités suisses (*Eveline Widmer-Schlumpf*, *Simonetta Sommaruga*), fréquence qui ne se retrouve pas dans les textes originaux.

- (2) C'est à cette fin que l'ancienne **ministre** des finances Eveline Widmer-Schlumpf a lancé la Troisième réforme de l'imposition des entreprises (RIE III). (textes influencés)

Enfin, la prédominance des lexèmes *affaires* et *étrangères* dans les deux sous-corpus nous a poussé à observer aussi les co-occurrences de l'expression *ministre des affaires étrangères*. Les listes obtenues, de tailles très différentes (Figure 11 et Figure 12), illustrent – presque à l'extrême – la tendance à l'international pour les textes originaux et au national pour les textes influencés.



Figure 11. Co-occurrences de Ministre des affaires étrangères dans les textes originaux



Figure 12. Co-occurrences de Ministre des Affaires étrangères dans les textes influencés

Il semble donc que les textes originaux s'inscrivent dans un contexte à la fois plus local (communal et cantonal) et international (rapport plus étroit à la France voisine et dimension mondiale avec une grande diversité de pays mentionnés).

Les textes influencés, en revanche, se concentreront plus sur la dimension administrative nationale face à l'étranger en tant qu'entité globale, le "non-suisse". Ces observations soulèvent de nouvelles questions, notamment quant à l'image de la Suisse et de ses régions (linguistiques, géographiques) véhiculée dans les textes influencés et originaux. L'étude des toponymes suisses mentionnés dans les deux sous-corpus permet alors de mieux comprendre le rapport entre la dimension locale et la dimension nationale.

4. Analyse des toponymes des sous-corpus

4.1 *Considérations méthodologiques*

Le balisage du corpus offre un autre indicateur de comparaison des sous-corpus "originaux" et "influencés": les noms de communes suisses. L'analyse des résultats requiert toutefois un filtrage manuel, qualitatif, pour les raisons suivantes:

- Certains toponymes étaient des homonymes de noms communs (par exemple *Premier*, commune du canton de Vaud et adjectif numéral ordinal).
- Certains toponymes présentent une graphie composée (par exemple *La Chaux-de-fonds*). Le corpus "lira" chaque unité non reliée par un tiret séparément, si bien qu'il faudra parfois rétablir l'intégralité du nom en question.

C'est pourquoi nous avons procédé au nettoyage des résultats selon les principes suivants:

- Toutes les prépositions et particules (sur, de) esseulées ont été supprimées, car elles ne fournissent pas d'information spécifique.
- Les toponymes divisés ont été reconstitués, en retenant uniquement la fréquence de la partie la plus spécifique du nom. Les décisions s'appuient sur une vérification au moyen des collocations et de la lecture des passages textuels correspondants.
- De même, les toponymes homographes de noms communs ou d'entités nommées autres (prénoms, noms de pays) ont été supprimés lorsque l'analyse des co-occurrences et des passages textuels correspondants ne montrait pas de prédominance de l'emploi toponymique.⁹

Enfin, notre analyse s'est limitée aux cinquante premières entrées de la liste filtrée, puis aux dix premières entrées pour une analyse plus en profondeur. Les

⁹ Il en a résulté l'exclusion des lexèmes suivants: *Sur, Brent, Les, Der, An, Le, La, Rue, Premier, Am, De, Watt, Mission, Chemin, Champagne*.

résultats suivants sont discutés directement à la lumière des co-occurrences des toponymes au sein des sous-corpus.

4.2 *Analyse des toponymes: résultats*

Les listes de noms de communes suisses obtenues dans le sous-corpus "originaux" et le sous-corpus "influencés" divergent sensiblement l'une de l'autre, comme le montrent la Figure 13 et la Figure 14. En effet, en comparant les deux cartes, on découvre un glissement vers les régions non francophones de Suisse dans les textes influencés.

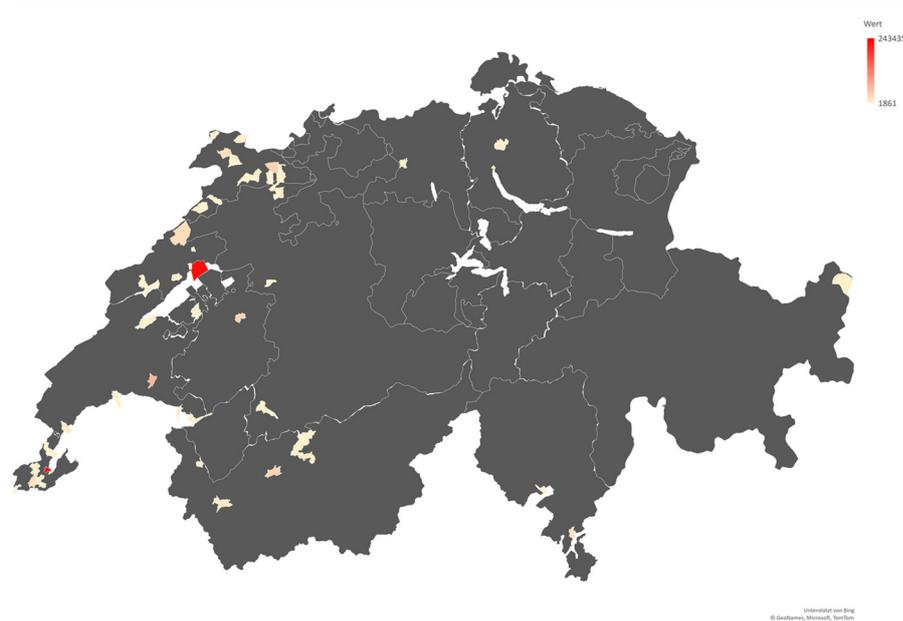


Figure 13. Toponymes présents dans les textes originaux (Swiss-AL). L'intensité de la couleur reflète la fréquence.

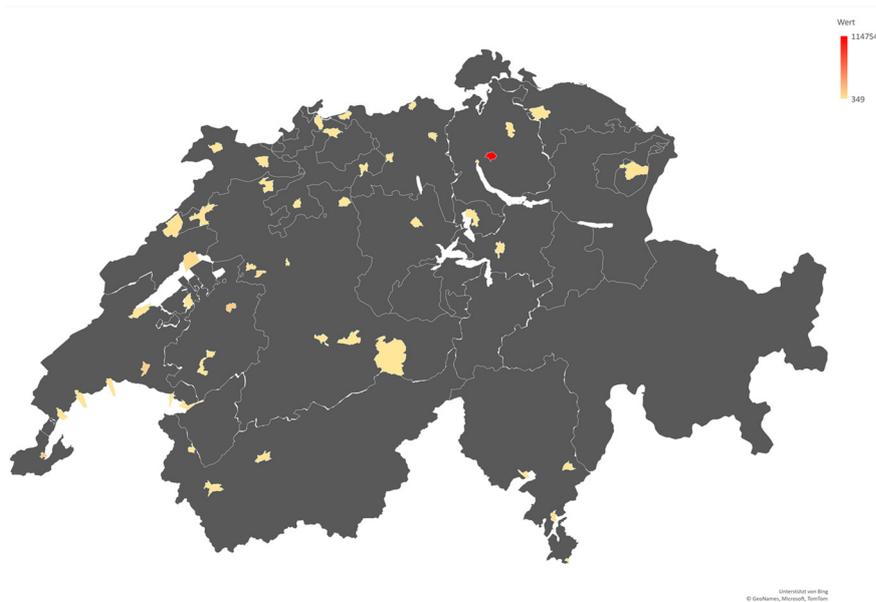


Figure 14. Toponymes présents dans les textes influencés (Swiss-AL). L'intensité de la couleur reflète la fréquence.

La première ville francophone mentionnée dans les textes influencés, Genève, n'arrive qu'en quatrième position après Dübendorf, Berne et Zurich. *Lugano*, en dernière position, indique une présence plus forte de la Suisse italienne dans les textes influencés.

Textes originaux		Textes influencés	
Ville	Fréquence	Ville	Fréquence
Genève	243435	Dübendorf	114754
Neuchâtel	230758	Bern/Berne	35846
Lausanne	62502	Zürich/Zurich	22195
Zürich/Zurich	59133	Genève	16503
Bern/Berne	55810	Fribourg	13211
Delémont	32455	Lausanne	11642
Sion	32274	Basel/Bâle	8682
Basel/Bâle	30910	Neuchâtel	5726
Fribourg	29789	Mühleberg	2011
La Chaux-de-Fonds	23575	Lugano	1749

Tableau 1. Toponymes les plus fréquents (Swiss-AL)¹⁰

¹⁰ La liste des textes influencés contenait également Appenzell et Schwyz (en 12^e et 13^e position). S'agissant uniquement de cantons et non de communes, ils n'ont pas été retenus.

4.2.1 Dübendorf

La présence de Dübendorf en tête de liste "influencés" (contre 300^e dans les textes originaux) peut soulever des interrogations. Ses co-occurrences sont similaires dans les deux sous-corpus: on y trouve essentiellement des fragments d'adresse (noms et numéros de rue, codes postaux), *aérodrome* (base militaire aérienne), *EAWAG* (l'Institut Fédéral Suisse des Sciences et Technologies de l'Eau) et *EMPA* (Laboratoire fédéral d'essai des matériaux et de recherche).

- (3) Il estime que **Dübendorf** pourrait être utilisé comme aéroport civil et décharger ainsi l'aéroport zurichois. (textes originaux)
- (4) Le 18 janvier dernier, l'EMPA (Eidgenössische Materialprüfungs- und Forschungsanstalt) de **Dübendorf** rendait un rapport cinglant [...] (textes originaux)
- (5) Projet UTF n o 362: Augmentation de la performance des stations d'épuration grâce à une meilleure sédimentation et au traitement biologique du phosphore dans le procédé SBR
Principaux partenaires: EAWAG, **Dübendorf** / STEP Jungholz, Uster [...] (textes influencés)

Il s'agit donc moins de textes à propos de la ville de Dübendorf que de publications produites par ou portant sur les acteurs nationaux qui y siègent. L'exemple de Dübendorf indique une référence beaucoup plus limitée aux instances de recherche fédérales dans les textes originaux.

4.2.2 Mühleberg

Autre commune germanophone de taille modeste dans la liste des textes influencés, Mühleberg semble surtout représenter la centrale nucléaire qu'elle abrite. Les co-occurrences des deux sous-corpus confirment cette première impression: *centrale*, *nucléaire*, *Beznau* sont les plus fréquentes. Toutefois, les textes originaux associent plus souvent *arrêt* (7^e position) et *fermeture* (11^e position) à Mühleberg, co-occurrences beaucoup moins fréquentes dans les textes multilingues.

- (6) L'entreprise table actuellement aussi sur l'horizon 2022 pour l'arrêt de **Mühleberg**. (textes originaux)

Les exemples de Mühleberg et Dübendorf montrent la complexité du décalage entre textes originaux et textes influencés: d'une part, les villes (et lieux) récurrents dans les corpus peuvent présenter des divergences inattendues et contre-intuitives. On s'éloigne donc d'une simple dichotomie local-national (suggérée par l'analyse des spécificités). D'autre part, les thématiques associées à ces villes et lieux peuvent varier selon le sous-corpus. Ces deux dimensions comparatives sont également porteuses pour les plus grandes métropoles.

4.2.3 Berne, Bâle, Zurich

Les grandes agglomérations de Zurich, Berne et Bâle sont plutôt caractéristiques des textes influencés (Tableau 1). Ces trois métropoles sont associées à des thématiques différentes selon les sous-corpus. Les co-occurrences des textes influencés sont essentiellement liées à des dénominations d'institutions et des adresses postales ou physiques (codes postaux, noms de rues).

- (7) Conseils et aide financière aux personnes retraitées: Pro Senectute dispose de services de consultation dans plusieurs grandes communes de Suisse Lavaterstrasse 60, Postfach, 8027 **Zürich** (textes influencés)

Zurich, Bâle et Berne se présentent donc, dans ce sous-corpus, surtout comme les lieux de production des textes influencés et/ou des lieux accueillant des acteurs producteurs de discours public. En revanche, les co-occurrences des textes originaux esquissent des associations différentes. Les co-occurrences de Zurich et de Bâle s'articulent notamment autour du sport (*FC, Lions*) et de l'économie (*ZKB, la Banque cantonale de Zurich*), ce qui n'est pas le cas dans les textes influencés.

- (8) Fribourg Gottéron a signé une performance de choix en s'imposant 3-2 face aux **Zurich** Lions, tenants du trophée, en 8e de finale de la Coupe de Suisse. (textes originaux)
- (9) Le SMI devrait continuer à évoluer légèrement en hausse à court terme [...] estiment lundi les chartistes de la Banque cantonale de **Zurich** (ZKB). (textes originaux)
- (10) [...] Servette, tout comme ses concurrents directs, espère une finale de Coupe de Suisse **Bâle** - Lucerne . (textes originaux)
- (11) Pour l'heure, Credit Suisse dispose de fonds propres nettement supérieurs au seuil de 7% [...]selon **Bâle** III de 16, 3 % à fin septembre. (textes originaux)

Pour Berne, en revanche, les co-occurrences des textes originaux se concentrent plutôt sur la notion de canton (première co-occurrence, suivie de divers autres cantons) et sur le domaine politique (*députés* en 19^e position et *socialistes* en 17^e).

4.2.4 Genève

Genève est la ville romande la plus mentionnée dans les textes originaux et influencés. Toutefois, on constate des différences dans les co-occurrences (Figure 15 et Figure 16) sur plusieurs niveaux.



Figure 15. Co-occurrences de Genève dans les textes originaux



Figure 16. Co-occurrences de Genève dans les textes influencés

Tout d'abord, les co-occurrences les plus fréquentes, *ville* pour les textes originaux et *canton* pour les textes influencés, dénotent bien un décalage d'échelon géo-administratif. Ce décalage se retrouve dans la présence accrue d'autres cantons (*Vaud*, *Jura*, *Valais*) et villes-cantons (*Zurich*, *Bâle*, *Neuchâtel*, *Bern*) dans les co-occurrences des textes influencés. La dimension cantonale est certes présente dans les co-occurrences des textes originaux (*Vaud*, *canton*), mais dans une bien moindre mesure.

- (12) Les cantons de Soleure, de Saint-Gall, des Grisons, d'Argovie, de Thurgovie, de Neuchâtel et de **Genève** peuvent, conformément à leur demande, proposer le vote électronique aux électeurs suisses [...] (textes influencés)
- (13) La DGM et la Ville de **Genève** soulignent la nécessité d'avoir des rabattements plus conséquents vers les haltes RER (textes originaux)

De plus, la co-occurrence *État* (pour l'État de Genève) figure également dans les textes originaux¹¹, et semble, dans ces usages, concurrencer *canton*. Le sous-emploi d'*État* dans les textes influencés suggère, à notre lecture, un décalage entre la vision fédérale de Genève et son image interne.

- (14) "Les autorités françaises ont unilatéralement décidé de reporter le début des opérations au mois de juin 2012", explique benoîtement l'Etat de **Genève** dans un courrier adressé le 23 février à l'Office fédéral de l'environnement. (textes originaux)
- (15) Vu l'existence d'une convention de location entre l'Etat de **Genève** et la Commune de Bernex pour une mise à disposition du local à sel [...] (textes originaux)

Les mentions de Genève sont donc plus focalisées à l'interne dans les textes originaux, et plus orientées sur la dimension nationale dans les textes

¹¹

On la trouve en 929^e position dans les textes influencés.

influencés. Les nombreuses co-occurrences des textes originaux liées à la vie culturelle locale (*Invitations, vos, Remportez, fêtes, Bibliothèque*) et à des formes d'action (verbes comme *réunis, vivre, augmente*) consolident cette impression.

- (16) One FM vous invite à Ciné du Lac à **Genève**. Remportez vos invitations VIP pour la Street Parade à Zürich. (textes originaux)
- (17) Besoin de conseils pour faciliter vos déplacements à **Genève** et alentours?(textes originaux)
- (18) Il y en a au moins trois: faire vivre la **Genève** internationale, la développer et veiller aussi à la rénovation de ses bâtiments. (textes originaux)

Ces observations se retrouvent dans les co-occurrences de Neuchâtel et Lausanne, non discutées ici pour des raisons de place. En revanche, Fribourg et les villes plus petites comme Delémont, Sion, présentent des co-occurrences similaires dans les deux sous-corpus, essentiellement liées à d'autres cantons (pour Fribourg) ou bien à une thématique sportive (*Gottéron* pour Fribourg, *FC* pour Sion) ou culturelle (*OSR*, Orchestre de la Suisse romande pour Delémont). Leur surreprésentation dans les textes originaux semble donc suivre la tendance observée pour Genève.

4.2.5 Lugano

La tendance observée pour les communes romandes s'inverse avec Lugano (10^e dans les textes influencés, 13^e dans les textes originaux). En effet, les co-occurrences des textes originaux sont essentiellement axées sur la politique (*congrès, PS, socialistes*) tandis que les co-occurrences des textes influencés comportent aussi des notions plus locales et géographiques, comme *lac*, première co-occurrence.

- (19) La lumière du soleil brille sur les vagues légères du **lac** de **Lugano**. (textes influencés)

Or, *lac* était présent dans les principales co-occurrences de Genève dans les textes originaux.

- (20) Plein d'idées pour une nouvelle version de la Rade Septante bureaux d'architectes ont soumis leurs projets pour le réaménagement des bords du **lac** à **Genève**. (textes originaux)

Ainsi est-il plus facilement question du lac de Genève dans les textes "romands", tandis que les textes fédéraux ou alémaniques¹² mentionnent plutôt le lac de Lugano. Cette différence induit un décalage entre l'image "originale" et l'image "influencée" associée à l'idée de lac. D'ailleurs, les lieux présents dans les co-occurrences de *lac* diffèrent selon le sous-corpus (Figure 17).

¹²

Nous rappelons ici que le corpus ne contient pas de textes italiens traduits vers le français.

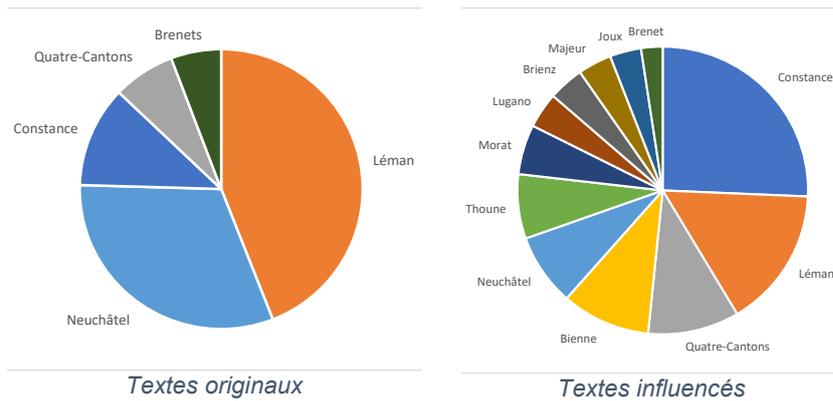


Figure 17. Lieux associés à lac (tirés des 20 premières co-occurrences)

5. Conclusions

D'un point de vue méthodologique, cette étude montre les limites de l'approche logométrique choisie face à la polysémie et à l'homonymie. En effet, il n'est pas possible d'exclure un ou plusieurs signifiés lors du décompte d'un signifiant, (par exemple *emploi, Etat/état etc.*). Il est donc utile de préciser des observations primaires de fréquence par une analyse des cooccurrences, une analyse textuelle qualitative ou encore des méthodes quantitatives automatisées de type *word embeddings*.

L'analyse des 301 sources du corpus Swiss-AL a mis en lumière une forte surreprésentation de l'allemand dans les textes écrits publiés en ligne en Suisse. Par ailleurs, l'analyse des sous-corpus ainsi générés a permis de déceler des tendances majeures. Tout d'abord, la représentation de la Suisse en tant qu'unité est plus forte dans les textes traduits ou similistraduits de l'allemand (influencés), de même que la représentation de l'étranger comme un ensemble généralisé. Ensuite, la représentation des niveaux décisionnels et actionnels locaux est plus forte dans les textes rédigés directement en français (originaux), et les pays étrangers sont plus souvent mentionnés, de manière individuelle (et moins comme un ensemble). Toutefois, le niveau local ne s'arrête pas aux frontières linguistiques: les textes originaux français traitent également (et fréquemment) de localités germanophones. Enfin, les thématiques associées aux différents lieux varient selon que les textes seront originaux ou influencés. Ainsi, si tous les textes mentionnent la même ville ou le même canton, l'accent ne sera pas forcément mis sur les mêmes éléments. Cette différenciation est aussi de rigueur pour les fonctions politiques (*président, ministre*) et probablement pour les noms géographiques (*lac*).

Ces trois observations posent la question des représentations mentales associées à certains noms (notamment géographiques) et même à l'identité suisse chez les Romands: différent-elles de celles des Alémaniques et des Tessinois? Et si c'est le cas, les organes de rédaction des textes influencés (traducteurs, rédacteurs non francophones en Suisse alémanique, etc.) en ont-ils conscience? En effet, on peut se demander dans quelle mesure les

divergences entre les perspectives identifiées dans les deux sous-corpus peuvent engendrer des distorsions dans la réception de textes et de discours. La question de l'influence de l'allemand sur le discours francophone nous paraît donc non seulement pertinente, mais, avec la généralisation du recours à la traduction automatique, de plus en plus brûlante. En effet, une meilleure connaissance des associations récurrentes dans les textes originaux français doit permettre aux auteurs non francophones d'opter pour une rédaction mieux adaptée (on parlera de localisation et de transcréation dans le domaine marketing, par exemple). La traduction automatique, instantanée et peu coûteuse, ne permet pourtant pas cette adaptation, et risque donc de nourrir encore longtemps l'impression que "Pour les Romands, c'est la même chose mais en français".

BIBLIOGRAPHIE

- Arquint, R. (2014): Plädoyer für eine gelebte Mehrsprachigkeit: Die Sprachen Im Räderwerk Der Politik in Der Mehrsprachigen Schweiz Und Im europäischen Ausland. Zurich (Verlag Neue Zürcher Zeitung).
- Chestermann, A. (2013): Models of what processes? Translation and Interpreting Studies 8(2), 155-168.
- Donzé, V. (2020, 01 21). La traduction hilarante d'une crème déride les Romands. www.lematin.ch (<https://www.lematin.ch/suisse/La-traduction-hilarante-d-une-creme-deride-les-Romands/story/20788063>)
- d'Urso, G. (2014). "L'italien est en difficulté, mais le français aussi". www.swissinfo.ch (https://www.swissinfo.ch/fre/plurilinguisme-en-suisse_-l-italien-est-en-difficult%C3%A9--mais-le-fran%C3%A7ais-aussi-/38242810)
- Krasselt, J., Dreesen, P., Fluor, M., Mahlow, C., Rothenhäusler, K. & Runte, M. (2020). Swiss-AL: A Multilingual Swiss Web Corpus for Applied Linguistics. Proceedings of the 12th Language Resources and Evaluation Conference, 4138-4144. Consultable à l'adresse: <https://www.aclweb.org/anthology/2020.lrec-1.509>
- Kübler, D., Kobelt, E. & Zwicky, R. (2020). Les langues du pouvoir. Le plurilinguisme dans l'administration fédérale. Lausanne: EPFL Press.
- Maingueneau, D. (1979). L'analyse du discours. Repères. Recherches en didactique du français langue maternelle, (51), 3-27.
- Manno, G. (2007): La situation du français en Suisse: considérations démolinguistiques et de politique linguistique. Constellations francophones, Publifarum, 7, Consultable à l'adresse: http://publifarum.farum.it/ezone_pdf.php?id=45
- Mayaffre, D. (2005). Analyse du discours politique et Logométrie: point de vue pratique et théorique. Langage et Société, 99-121.
- Mayaffre, D. (2007). Analyses logométriques et rhétorique du discours. In S. Olivési (éd.), Introduction à la recherche en SIC. Grenoble: Presses Universitaires de Grenoble, 153-180.
- Office Fédéral de la Statistique. (2020). Langues Principales en Suisse. Population Résidente permanente
- SRF (10. 03 2016). Wie bilingue ist die Schweiz? www.srf.ch. (<https://www.srf.ch/news/schweiz/wie-bilingue-ist-die-schweiz>)

- Stoffel, M. (2019). Mehrsprachige Gesetzesredaktion. Mémoire de Master. Winterthur.
- Stücheli-Herlach, P., Ehrensberger-Dow, M. & Dreesen, Ph. (2018): Energiediskurse in der Schweiz: Anwendungsorientierte Erforschung eines mehrsprachigen Kommunikationsfelds mittels digitaler Daten. Working Papers in Applied Linguistics 16. Winterthur (ZHAW Zürcher Hochschule für Angewandte Wissenschaften). Consultable à l'adresse: <https://doi.org/10.21256/zhaw-7106>
- von Busekist, A. (2007): Cannibales et gourmets: Quelques recettes d'équilibre linguistique. In M. Werner (éd.), Politiques et usages de la langue en Europe. Paris (Maison des sciences de l'homme) 101-119.