

## Les effets de la distance sur le contenu d'un cours : Une analyse avec LSA

*Philippe Dessus*

### Résumé

---

Nous réexaminons ici les données d'une précédente expérimentation comparant le discours d'un enseignant à distance avec le discours du même enseignant, en présence de ses étudiants. Nous avons précédemment montré que ces deux discours ne différaient pas significativement sur les plans lexical, morpho-syntaxique et pragmatique. Cette nouvelle étude utilise l'analyse de la sémantique latente pour ré-analyser les deux modalités du discours. Les résultats ne révèlent toujours aucune différence, mais les capacités de LSA à déterminer que deux séquences sont sémantiquement très proches, malgré des différences dans le texte, peuvent être utilisées dans la conception d'un système d'enseignement à distance que nous détaillons. Cette méthode permet également de mesurer la variable « structure du cours » issue de la théorie de la distance de transaction.

### Abstract

---

In this article we analyze the data of an earlier experiment in which two modes of a teachers' discourse were compared: at a distance versus face-to-face. The results showed no significant differences between the two forms of discourse concerning its lexical, morphosyntactical, and pragmatological forms. We present here a new analysis in which the Latent Semantic Analysis method is used to reanalyze the data in a semantic perspective. Although there was again no significant difference between the two modes, we show the interesting capability of LSA to match semantically two separate pieces of text. We then discuss how LSA could be used first, as the heart of a distance learning system and second, to operationalize the structure variable of transactional distance theory.

---

Il est souvent implicitement posé que le fait d'enseigner à distance par l'intermédiaire d'un outil crée une différence avec l'enseignement dans la même situation, mais sans outil et en présence. Sommairement, deux issues peuvent être considérées, respectivement tenues par Clark (1983) et par Kozma (1991) : considérer que l'outil n'a aucun effet (respectivement, a un effet) sur les principales variables de l'enseignement ou de l'apprentissage. Nous ne détaillerons pas ici ce débat, mais essaierons d'évaluer l'effet de la distance (à dispositif et méthode pédagogiques équivalents) sur le contenu du cours, évalué par une méthode d'analyse textuelle

statistique. Il ne s'agit pas de dire, *a priori*, que le média en général n'exerce aucun effet sur l'apprentissage ou l'enseignement, mais plutôt de déterminer quelle partie du média, dans quel contexte et surtout sur quelles variables il pourrait en avoir un.

Les études que nous avons réalisées dans ce domaine (Dessus et Lemaire, 1999) ont abouti à un constat de non différence significative entre groupe expérimental et groupe contrôle, fréquent dans la littérature (Russell, 2002). Pour autant, il ne faut pas conclure, comme c'est souvent le cas, par une *égalité* de performances entre les deux groupes (Lockee, Burton, et Cross, 1999) : les hypothèses formulées, ainsi que les tests statistiques utilisés, permettent de mettre au jour une éventuelle différence entre groupes (*i.e.*, de rejeter une hypothèse nulle d'égalité), mais pas d'établir une quelconque égalité entre groupes. Certains chercheurs en enseignement à distance viennent un peu rapidement, et de manière erronée, à dire que les effets de l'enseignement à distance sont identiques à ceux de l'enseignement en présence seulement parce que des différences non significatives entre les performances de deux groupes de sujets ont été observées. Pourtant, la variable observée peut ne pas être pertinente pour la recherche. Il faut donc déterminer des systèmes, des situations et des variables dans lesquels des différences s'exprimeront. La spécification de ces trois ensembles de paramètres est également essentielle pour un dispositif d'enseignement à distance : trop souvent, en effet, les auteurs d'un système supposent qu'il est efficace sans en apporter de preuve.

L'objet de cet article est de ré-analyser les données d'une recherche qui a consisté à comparer le discours d'un enseignant à distance à celui du même enseignant en présence. Cette comparaison n'avait pas abouti à des différences significatives, nous proposons ici une analyse d'une nouvelle variable, effectuée avec l'aide d'une méthode statistique d'analyse de grands corpus textuels. Commençons par détailler la littérature dans ce domaine.

### Enseignement à distance et analyse du discours

La recherche expérimentale sur l'enseignement à distance (Dessus, Lemaire, et Baillé, 1997; Phipps et Merisotis, 1999) fait l'objet de critiques, liées en partie au problème dont nous avons fait état dans l'introduction : les variables sélectionnées pour en rendre compte ont majoritairement concerné des aspects de surface (principalement les caractéristiques des élèves ou des enseignants). Un deuxième problème est celui lié à la rareté des théories dans le domaine : les résultats d'études sur l'enseignement à distance sont d'une part rarement reliés à des théories les prédisant et les expliquant et, d'autre part, les théories existantes sont rarement vérifiées empiriquement. Nous essayons ici de pallier ces problèmes avec une opérationnalisation de la théorie de la distance de transaction.

De nombreuses recherches ont analysé le discours de l'enseignant au sein d'un dispositif d'apprentissage collaboratif (Baker, de Vries, Lund, et Quignard, 2001), mais il en existe beaucoup moins qui ont pour objet des systèmes d'enseignement à distance « classiques » (Lemaire, Marquet, et Baillé, 1996). Pourtant, il est possible de considérer le cours de l'enseignant comme un discours et d'utiliser les outils d'analyse du discours pour détailler ses caractéristiques. Le fait que l'enseignant soit distant physiquement des étudiants à qui il s'adresse peut causer des différences dans son discours, mais on a montré depuis longtemps que, d'une part, on peut considérer qu'il existe une distance symbolique entre l'enseignant et ses élèves dans un enseignement en présence et que, d'autre part, il était difficile de justifier théoriquement *sur quelles variables* cette distance avait un effet (la communication, le contenu, l'apprentissage, etc.). La théorie de la distance de transaction (Moore, 1993), que nous présentons maintenant, nous permet justement de prédire certains effets.

### *Une opérationnalisation de la distance de transaction*

La distance de transaction est une distance fictive, non physique, entre l'enseignant et l'élève, variant selon deux paramètres principaux : – *la structure du cours* : « La structure rend compte de la rigidité ou de la flexibilité des objectifs éducatifs du programme, des stratégies d'enseignement et des méthodes d'évaluation » (Moore, 1993, p. 26) (*i.e.*, ses possibilités d'adaptation aux élèves) et – *le dialogue* (*i.e.*, la richesse des interactions entre enseignants et élèves, ou interélèves). La distance de transaction d'une situation d'enseignement est d'autant plus grande qu'on minimise le dialogue tout en rigidifiant la structure du cours. Cette notion est intéressante en ce qu'elle permet de rendre compte de la perception de distance par les étudiants et l'enseignant dans des situations d'enseignement. Toutefois, elle a été critiquée pour son manque de précision et la difficulté de pouvoir la tester (Garrison, 2000) :

- les exemples donnés par Moore ne concernent que les effets de médias (matériel), et non de situations. Or, la flexibilité des objectifs et méthodes dépendent de l'enseignant et sont faiblement contraints par les médias;
- la variable « structure » est difficile à déterminer quantitativement : les exemples donnés par Moore de programmes très structurés (*i.e.*, programme diffusé par vidéocassette) prêtent à discussion, car il semblent plutôt être des exemples de cas où la variable « dialogue » est inexistante;
- enfin, la variable « structure » est difficile à mesurer dans l'absolu : elle ne peut se mesurer qu'en comparant deux types de discours d'enseignant pendant son cours. Des différences entre ces deux types

montreraient un cours peu structuré, alors que de faibles différences montreraient un cours très structuré.

Malgré quelques tentatives (Chen, 1998; Saba et Shearer, 1994), il existe peu de validations empiriques de cette notion. La raison en est sans doute la difficulté de mesurer empiriquement cette distance, la plupart du temps issue des avis des étudiants. Ces études se centrent principalement sur des aspects sociaux et de communication. Il serait intéressant de vérifier l'impact de cette notion sur un aspect important du cours, son contenu. Il nous faut pour cela une méthode valide de mesure du contenu : nous posons que le contenu du cours sera d'autant plus structuré que les deux types de discours (en présence et à distance) seront évalués comme semblables par une méthode d'analyse statistique que nous présentons maintenant.

### *Présentation de l'analyse de la sémantique latente*

L'Analyse de la sémantique latente (désormais LSA, pour *Latent Semantic Analysis*) commence à être utilisée pour rendre compte du contenu des discours ou textes dans certains systèmes d'enseignement à distance. Dans cette partie, nous détaillons son principe de fonctionnement, puis recensons quelques utilisations. Le principe de LSA est le suivant : le sens d'un mot est défini statistiquement à partir de l'ensemble des contextes (*i.e.*, paragraphes, phrases, textes) dans lesquels ce mot apparaît. Par exemple, le mot *avion* va apparaître souvent conjointement à des mots comme *décoller*, *ail*, *aéroport* et très peu fréquemment à des mots comme *pédagogie* ou *dinosaure*. Cependant, cette information statistique sur le contexte d'un mot n'est pas suffisante pour en définir le sens, puisqu'elle ne dit rien quant aux liens sémantiques avec tous les autres mots n'apparaissant jamais conjointement à ce mot. Il faut pour cela un mécanisme permettant de croiser les informations de co-occurrence propres à chaque mot. Ensuite, les différents mots du corpus sont projetés dans un espace d'environ 300 dimensions, ce qui autorise leur comparaison et, par sommation des vecteurs des mots, la comparaison de paragraphes.

De très nombreux résultats expérimentaux montrent que LSA est un modèle cognitif acceptable de la représentation et l'acquisition de connaissances à partir de textes (Lemaire et Dessus, à paraître). Par exemple, on a montré que LSA pouvait noter, avec des performances proches de juges humains, la qualité de résumés de textes en les comparant avec leurs textes sources (Foltz, 1996). LSA traite les mots par « paquets » pour évaluer la connaissance. Cette méthode comporte certains avantages, dont l'indépendance du domaine et la rapidité du traitement, mais aussi des inconvénients, comme l'absence de la prise en compte des aspects syntaxiques du discours. Détaillons maintenant deux systèmes utilisant LSA.

## *Systèmes d'enseignement et d'apprentissage à distance basés sur LSA*

CUMULI (Reffay et Chanier 2001) est un système d'enseignement à distance permettant de faciliter les communications asynchrones entre apprenants travaillant en groupe dans le même domaine, via un forum de discussion. La vitesse de diffusion d'une question aux membres d'un groupe dépend ainsi de certains paramètres, tels leur niveau d'intérêt, de compétence et de réactivité. LSA permet d'estimer ces niveaux, à partir des réponses de chaque apprenant et, ainsi, chaque question est envoyée aux participants par vagues : les plus faibles, les moyens et éventuellement les forts, tant qu'aucune réponse satisfaisante n'a été apportée à la question.

*Knowledge Forum* (Scardamalia et Bereiter, 1994) est un système de construction collaborative de connaissances qui consiste en un forum de discussion et un site sur lequel placer des documents hypermédias. Les participants donnent leur point de vue sur des questions ouvertes, commentent et critiquent celui des autres, et construisent ainsi des connaissances à propos du domaine discuté. Les différents appariements entre connaissances et discussions se faisant par les apprenants, des chercheurs (Teplovs et Van Aalst, 1998) ont utilisé LSA pour les réaliser automatiquement. Il s'agit de déterminer le statut des productions d'élèves (*i.e.*, explication, découverte, résultat, question, expression d'un doute) en comparant ces dernières à des documents de référence.

Notre propre travail utilise également LSA pour déterminer de quelle manière deux types de discours d'enseignant (en présence *versus* à distance) sont semblables. Contrairement aux systèmes cités plus haut, il se centre donc plus sur l'enseignant, en supposant que l'analyse de son discours est un moyen fiable de rendre compte du contenu traité dans le cours. De ce fait, et surtout du fait de la situation observée (une vidéoconférence), il est peu question des interactions enseignant-élève, rares dans ce type de situations.

### Comparaison du discours d'un enseignant en présence versus à distance

Nous étudions quelques variables du discours d'un enseignant faisant un cours par audioconférence (guidé par un diaporama), et les comparons selon qu'il s'adresse à ses étudiants à distance ou en présence. Cette étude reprend et ré-analyse les données recueillies dans une expérimentation précédente, dans laquelle le discours de l'enseignant était comparé selon que ce dernier délivrait son cours en présence, dans un amphithéâtre ou à distance, via audioconférence. Aucune différence significative intersituation n'avait été mise au jour, ce qui nous amène maintenant à prendre en compte les aspects sémantiques du discours, en utilisant LSA.

### *Description de l'expérimentation*

*Contexte.* Une première expérimentation a déjà visé à mesurer les effets de la distance sur le discours de l'enseignant (Lemaire, Dessus, et Baillé, 1998). Les aspects du discours pris en compte étaient les niveaux lexical, morphosyntaxique et pragmatique. Nous allons ici ré-exposer la procédure de l'expérimentation, en laissant de côté ses résultats, déjà discutés ailleurs (Dessus et Lemaire, 1999). Elle s'appuie sur un cours de comptabilité nationale de première année de DEUG de Sciences économiques (première année universitaire), dispensé de deux façons, de manière à ce que la seule variable susceptible d'avoir un effet sur le discours de l'enseignant soit la distance : – *en amphithéâtre*, sous forme de cours magistraux accompagnés de transparents, suivis par 109 étudiants; – *par un cours à distance* (audioconférence), en direct, appuyé par les mêmes transparents. Ce cours est suivi par 57 étudiants. L'enseignant n'est pas visible par les étudiants. Un assistant, sur place, est chargé de donner la parole aux étudiants.

*Résultats de la première analyse de données.* Une première série d'analyses a été réalisée en comparant le discours de l'enseignant sur les plans lexical, morphosyntaxique et pragmatique. Décrivons rapidement en quoi elle a consisté :

- l'analyse lexicale s'est fondée sur des travaux en lexicométrie (Muller, 1992). Des indicateurs de richesse et distribution lexicale ont été calculés;
- l'analyse morphosyntaxique a utilisé le modèle de Bronckart (1985), qui caractérise, à partir de l'analyse d'unités lexicales, la condition de production d'un discours en la référant à des situations architypiques (narration, en situation, théorique);
- l'analyse pragmatique (Henri et Ricciardi-Rigault, 1996; Searle, 1982), permet de classer les actes de langage mis en œuvre dans une situation de communication en cinq catégories : assertif, directif, expressif, engageant et déclaratif.

Aucune différence significative entre les deux situations n'a été relevée, ce qui ne nous autorise bien sûr pas à conclure que les discours *sont les mêmes*. Il convient seulement de trouver une méthode d'analyse qui fasse apparaître des différences. L'analyse du contenu, du point de vue du sens du discours n'ayant été faite que partiellement (seulement via l'analyse pragmatique), nous avons décidé de reprendre l'analyse de ce point de vue.

*Problématique et hypothèse de la deuxième analyse.* On peut essayer de mesurer dans quelle mesure la distance de transaction dans les deux situations observées est différente. Il nous faut pour cela étudier le dialogue et la structure du cours. Le dialogue enseignant-étudiants est une donnée aisée à observer. Un comptage des épisodes de dialogue, dans les

deux situations, montre que l'enseignant à distance a dialogué 4 fois avec les étudiants, contre deux fois en présence. Le contenu du dialogue portait soit sur la compréhension des étudiants, soit sur le traitement de tel ou tel contenu du cours, ce qui montre que, dans la pratique, il est déjà peu évident de distinguer les deux principaux paramètres jouant sur la distance de transaction. En revanche, la structure du cours, bien que contrainte, dans les deux modalités, par le contenu des diapositives projetées aux étudiants, peut être sujette à plus de variations. Ainsi, nous faisons l'hypothèse que c'est la structure qui va varier dans les deux modalités. Cette variation va pouvoir avoir un effet, d'un point de vue sémantique, sur le discours de l'enseignant; et cette variation pourra être mise en évidence par une méthode d'analyse statistique de grands corpus de données.

*Traitement des données pour la deuxième analyse.* Le discours de l'enseignant (corpus de 26 000 mots, 65 diapositives, dont 62 commentées dans les deux situations) a été segmenté par diapositive et traité par LSA, conjointement à un corpus de 10 millions de caractères provenant du *Monde*, corpus qui permet d'affiner la connaissance de la langue. Ensuite, LSA détermine la proximité sémantique entre le discours de l'enseignant pendant l'exposition d'une diapositive en présence et le discours à propos de la même diapositive à distance (valeurs comprises entre - 1 et 1). Comme la taille de la séquence de chaque diapositive est une variable qui joue de manière importante sur les résultats, nous avons supprimé de l'analyse les séquences de taille inférieure à 100 mots (valeur arbitraire, également choisie par d'autres (Teplovs et Van Aalst, 1998), dans au moins une des modalités. Nous sélectionnons de cet ensemble de paires de séquences celles s'écartant de plus d'un écart type de la moyenne des proximités. Ensuite, l'enseignant ayant donné le cours a déterminé les différences de contenu, pour chacune des paires de séquences sélectionnées. Après avoir supprimé des séquences ces différences (représentées en italiques dans les tableaux ci-dessous), nous avons recalculé la proximité sémantique entre paires, afin de vérifier si cette dernière augmentait ou baissait. Bien évidemment, lorsque le nombre de mots restant était inférieur à 100 pour l'une des séquences, la paire était supprimée de l'analyse.

*Résultats.* La première valeur calculée est la proximité moyenne entre les séquences en présence et celles à distance : 0,80 (écart type 0,08). Cette valeur, très élevée, est très supérieure à la valeur moyenne de la proximité de deux diapositives en présence (0,06) ou à distance (0,07). Il y a donc une différence très faible entre les deux types de discours. Ensuite, nous avons vérifié deux types de comparaisons : – des discours évalués comme très proches par LSA (*i.e.*, dont la proximité est supérieure à un écart type au-dessus de la moyenne, soit 0,89), bien que différents formellement; – des discours évalués comme différents par LSA (*i.e.*, dont la proximité est

inférieure à un écart type au-dessous de la moyenne, soit inférieure à 0,73). Concernant le premier point, notons que LSA peut considérer comme très proches des séquences parfois formellement différentes (*voir tableau 1 ci-dessous*) :

- explication d'une diapositive. Lorsqu'un exemple est donné dans une situation et non dans l'autre, la proximité reste tout de même importante (*i.e.*, pour la diapositive 7, la proximité interséquence est de 0,89, malgré des différences importantes, dont la suppression est sans effet sur cette valeur).
- détail d'un tableau, ou d'une ou plusieurs équation(s). Le caractère mathématique (équation) ou systématique (tableau) d'une diapositive

Tableau 1

Discours de la séquence de la diapositive 7 (Explication des classements des UP). *La proximité entre les deux discours est élevée (0,89), malgré des différences de structure importantes (en italique).*

Discours en présence	Discours à distance
<p>7. Alors, je reviens sur les classements. Alors, je vais classer selon la production principale et je vais classer les unités institutionnelles selon la production principale, j'obtiendrai donc des secteurs. Donc un secteur productif, c'est un regroupement d'unités institutionnelles, qui produisent le même ensemble de produits, le même groupe de produits, d'accord? Si je classe par contre selon les produits, je classe, je regroupe des unités de production homogène et j'obtiens des branches. Donc une branche productive est un regroupement des unités de production homogène. Le secteur productif est le regroupement d'unités institutionnelles, par rapport à leur production principale, bien sûr, mais c'est un regroupement d'unités institutionnelles. Donc deux critères différents.</p>	<p>7. Alors, revenons maintenant sur les classements des unités de production. Alors je peux classer selon la production principale, et à ce moment-là j'obtiens un regroupement par secteurs. Donc, les secteurs regroupent des unités institutionnelles, selon leur fonction, leur production principale. <i>Bon, je vais retrouver les producteurs de... de savon, les producteurs d'huile, les producteurs de l'automobile. Mais, dans le secteur, j'aurai une activité principale et des activités secondaires, éventuellement, prenons, je sais pas un producteur de yaourt, il aura peut être comme sous-produit ou comme activité secondaire de produire je sais pas moi, du lait, ou du beurre, si il fait des yaourts allégés, il récupérera le gras et il en fera du beurre, bien;</i> par contre, deuxième type de regroupement, si je regroupe selon les unités de production homogène, donc selon la logique des produits, je vais obtenir des branches, donc une branche ne produit qu'un élément de la nomenclature, donc une branche un produit, plus exactement, une branche, un groupe de produits. Vous voyez que la logique est un peu, est complètement différente, les secteurs regroupent des unités institutionnelles, les branches regroupent des unités de production homogène. Un secteur a plusieurs types de productions, une branche n'a qu'un type de production, type de production qui, bien entendu, dépend du niveau de regroupement de la nomenclature. Si on a une nomenclature avec quelques milliers de rubriques, c'est pas la même chose que si on a une nomenclature avec quarante rubriques.</p>



rend souvent les discours très proches, même si des exemples sont ajoutés dans une modalité.

Nous pouvons constater que LSA est une méthode d'analyse du contenu du discours assez robuste : des séquences à propos d'un même contenu ont été évaluées comme semblables, même si des différences de formulation, voire d'exemples, ont été introduites par l'enseignant. De plus, il est à noter que les exemples sont la plupart du temps énoncés dans la situation à distance.

Voyons enfin ce qu'il en est des séquences évaluées comme différentes selon le seuil fixé. Dans ce cas, les valeurs faibles correspondent à des séquences dont les exemples ou les digressions donnés dans une modalité étaient sensiblement différents de ceux de l'autre modalité de la même séquence. Dans ce cas, il importe de vérifier si, lorsque les différences entre discours sont supprimées, la valeur de proximité entre les deux séquences augmente ou baisse. Nous voyons que ce n'est pas toujours le cas. Si la valeur de proximité entre les séquences « distance » *versus* « en présence » de la diapositive 4 (voir tableau 2 ci-dessous) augmente bien dans le sens attendu (de 0,63 à 0,68) quand on supprime les exemples, indiqués en italiques, ce n'est pas le cas des autres exemples (*e.g.*, la suppression des exemples et digressions de la diapositive 5, voir tableau 3 ci-dessous, est sans effet sur la proximité interséquence, qui est de 0,67). Ce phénomène peut s'expliquer ainsi : les mots de l'exemple cité dans une modalité, même s'ils ne sont pas retraduits dans l'autre, jouent un rôle dans le calcul de la proximité (*i.e.*, ils parlent de production, d'activité, de branche). Les supprimer purement et simplement peut donc, parfois, faire baisser les valeurs de proximité.

#### Discussion : Applications à un système d'enseignement à distance

Décrivons maintenant quelques fonctionnalités d'un système d'enseignement à distance qui tire parti des résultats ci-dessus. Nous avons montré que LSA pouvait analyser automatiquement le contenu d'un cours délivré par un enseignant et que cette analyse était suffisamment fiable pour – discriminer les cas où des digressions à la marge du contenu survenaient; – ne pas discriminer les cas où étaient donnés des exemples supplémentaires liés au domaine du cours. La plupart des systèmes d'enseignement à distance (*i.e.*, *WebCT*, *BSCW*, les MOOS) comportent au moins deux fonctionnalités indépendantes utilisées de manière conjointe par les étudiants : un cours sous forme hypertextuelle et un forum de discussion, généralement asynchrone. Les étudiants peuvent prendre connaissance du cours, en naviguant dans les différentes pages, puis discuter de points précis du cours par le forum de discussion. L'intégration d'un outil tel que LSA pourrait se faire à ces deux niveaux.

Tableau 2

Discours de la séquence de la diapositive 4 (branches/secteurs). La proximité entre les deux discours passe de 0,63 à 0,68 quand on supprime les différences (en italique).

Discours en présence	Discours à distance
<p>4. Alors, branches/ secteurs, j'ai jusqu'à présent, essentiellement, j'ai uniquement d'unité institutionnelle, unités institutionnelles qui se définissent par une autonomie de décision. Ces unités institutionnelles ont une production principale, des productions secondaires. Ah, faites bien attention que cette fois je ne parle plus de fonctions, mais de productions, ce qui va me renvoyer à des produits, que je définirai plus tard, et cette unité institutionnelle pourra être composée de plusieurs unités de production homogène. C'est la production qui est homogène, c'est pas les unités. Donc, ces unités de production homogène, qui ne produiront qu'un produit, de ma nomenclature.</p>	<p>4. Alors, commençons donc par branches et secteurs. Alors, ici, il y a un problème aujourd'hui, voilà, donc unité institutionnelle, donc je rappelle que les unités institutionnelles, l'unité institutionnelle qui m'intéresse ici, c'est une unité institutionnelle qui produit. <i>Donc, qui n'est pas un ménage, qui consomme. Mais, ça peut être un ménage au sens d'entreprise individuelle. Donc, je vais retrouver mes sociétés, mes entreprises d'assurance, mes [actions?] financières, mes APU mes APR.</i> Et, du point de vue de la production, une unité institutionnelle a une production principale et des productions secondaires. <i>Prenez, je sais pas, toujours mes, fana... On parle toujours de l'industrie automobile, mais bon, vous savez que c'est l'enfant chéri des économistes. Bon, une entreprise automobile quelconque comme Peugeot produit d'abord de l'automobile et puis secondairement de l'outillage; ils produisaient à une époque du vélo, je sais pas s'ils en font toujours, c'étaient des productions secondaires. On peut avoir aussi des entreprises comme Renault qui produit prioritairement de l'automobile, secondairement du camion.</i> <i>Bien.</i> Alors, donc, deux types de production, et l'unité de production homogène, l'UPH, va se référer donc, va être un découpage, si on veut l'unité institutionnelle, qui permet de n'avoir, dans l'UPH, que un type de production. Donc on aura une liaison étroite entre l'unité de production homogène et production, par rapport évidemment à la nomenclature, on va y revenir.</p>

Tout d'abord en intégrant LSA en tant qu'outil de recherche d'informations dans un cours hypertextuel, et éventuellement le forum de discussion. Les capacités de LSA en tant qu'outil de recherche d'informations peuvent permettre de développer une certaine *sensibilité aux connaissances* (Ogata et Yano, 2000). Il s'agit de favoriser la collaboration entre étudiants à l'aide d'un outil d'analyse automatique de connaissances. Dans ce cadre le logiciel adresse automatiquement des messages tels que « X est en train de lire le même type de connaissances que vous êtes en train de lire » ou « Y et Z sont en train de discuter sur le même genre de thème que celui qui vous intéresse ». Cette fonctionnalité pourrait de plus être combinée à celle que nous avons mise au point, toujours à partir de LSA (Dessus, 2002) : l'extension de ce mécanisme de recherche d'informations à la recherche

Tableau 3

Discours de la séquence de la diapositive 5 (Les critères). La proximité entre les deux discours (0,67) ne varie pas quand on supprime les digressions.

Discours en présence	Discours à distance
<p>5. Alors, si je regarde donc les critères, que je vais utiliser par rapport à mes unités. Je pourrais d'abord utiliser un critère institutionnel et un critère fonctionnel. Le critère institutionnel, je le rappelle, c'est l'autonomie de décision de la fonction économique principale. Donc, les unités institutionnelles sont des centres de décision. Et à travers leur ..., on essaie de trouver un certain nombre d'indicateurs qui nous permettent de préciser les centres de décision. Le critère institutionnel, l'application du critère institutionnel concerne tous les agents, tous les agents présents sur l'espace économique national vont être répertoriés, classés grâce au critère institutionnel. Le critère fonctionnel nous renvoie au produit. Bon j'en parlerai tout à l'heure mais vous voyez bien que le produit, ça nous renvoie aux nomenclatures de produits dont j'avais parlé en statistique. Donc, et les unités, le critère fonctionnel, l'unité de production homogène, les unités de production homogène, ce critère ne s'appliquera qu'à un certain nombre d'agents qui sont les agents producteurs, donc j'exclus tous les agents qui sont purement consommateurs. Donc les ménages, en tant qu'agents consommateurs, ne participent pas des unités de production homogène.</p>	<p>5. Alors, les critères, les critères de définition de l'unité institutionnelle, critères de définition des UPH. Pour les unités institutionnelles, j'ai bien entendu le critère institutionnel, le centre autonome de décision et ce critère concerne tous les agents, donc j'ai quelque chose de général, ici, si vous voulez, j'ai un, je place l'ensemble de mes unités. Par contre, pour les unités de production homogène, c'est l'unité de production d'un produit, et ne concerne bien entendu que les agents qui produisent. Donc, ceux qui ne produisent pas sont exclus, ils ne sont pas des unités de production homogène. J'ai donc une séparation, dans le total de mes agents producteurs et je vais m'intéresser, pour les unités de production homogène, qu'à un seul type d'agent, ceux qui produisent des biens ou des services. Donc voyez que ça nous renvoie toujours à cet élément de définition de la production. <i>Et dans l'ancien système de comptabilité nationale les administrations publiques, ne produisant rien, ne pouvaient pas être classées dans les unités de production homogène, on verra que ça a, ça a quelques effets sur la façon dont on analyse la suite.</i></p>

sur Internet d'articles de référence à propos d'un domaine, à partir d'un état de la compréhension des étudiants.

Ensuite, en intégrant LSA en tant qu'assistant dans un forum de discussion. Un des problèmes, souvent soulevé, des forums de discussion est celui de son arbitrage (*i.e.*, la supervision des discussions par un enseignant, un tuteur). La charge cognitive de cette activité d'arbitrage croît avec le nombre d'étudiants en ligne, et nécessite donc, soit de nombreux enseignants, soit des outils d'aide à l'arbitrage. LSA pourrait servir de moteur à un tel outil, dans la lignée de ce que proposent certains (voir § Systèmes d'enseignement et d'apprentissage basés sur LSA). Ses avantages sont la rapidité de l'analyse et l'indépendance du domaine. Ainsi, les deux fonctionnalités (cours et forum) ne seraient pas indépendantes : le ou les cours

sont traités par LSA, qui utilise les connaissances issues de ces derniers pour analyser les discussions des forums.

Cet article a eu pour but d'opérationnaliser la notion de structure de cours, à partir de la mesure de la proximité sémantique des séquences de discours d'un enseignant, selon deux modalités, à distance *versus* en présence. Les résultats confirment les précédents, obtenus sur d'autres aspects du discours : les différences entre les deux types de discours sont très faibles, ce qui montre que le discours de l'enseignant, pour ce cours, est très structuré et que le fait d'être à distance n'implique pas une différence suffisamment importante dans le contenu du cours pour qu'on puisse noter une différence dans sa structure. Par conséquent, nous ne sommes pas arrivés à montrer que la *distance en tant que telle* a un effet sur le discours, ni dans l'étude précédente, ni dans celle-ci. Nous ne pouvons pas pour autant conclure que les deux situations sont identiques. Il reste encore à trouver des variables dépendantes pertinentes qui permettront la mise au jour de différences. Toutefois, l'utilisation de LSA à des fins d'analyse du contenu de cours a permis de montrer qu'il pouvait être un outil intéressant. En effet, des séquences de discours ont été jugées semblables à d'autres, à propos du même contenu mais avec des exemples différents, bien que toujours liés à ce contenu.

### Remerciements

Nous remercions vivement Pierre Bailly, qui nous a permis de réaliser une observation de son cours; ainsi que Benoît Lemaire, Pascal Marquet et les participants du séminaire du Laboratoire des sciences de l'éducation de Grenoble, pour leurs commentaires d'une version précédente de cet article.

### Références

- Baker, M.J., de Vries, E., Lund, K., et Quignard, M. (2001). Interactions épistémiques médiatisées par ordinateur pour l'apprentissage des sciences : bilan de recherches. *Sciences et Techniques Éducatives*, 8, 21-32.
- Bronckart, J.-P. (Éd.). (1985). *Le fonctionnement des discours*. Neuchâtel : Delachaux et Niestlé.
- Chen, Y.-J., & Willits, F.K. (1998). A path analysis of the concepts in Moore's transactional distance theory in a videoconferencing learning environment. *Journal of Distance Education*, 13(2), 51-65
- Clark, R.E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53, 445-459.
- Dessus, P. (2002). *Simulating student comprehension with LSA to deliver distance course readings*. Poster presented at the ESSCS Workshop on Multidisciplinary Aspects of Learning. Clichy : ESSCS.
- Dessus, P., et Lemaire, B. (1999). La technologie des médias joue-t-elle un rôle dans la transmission des connaissances? In S. Agostinelli (Ed.), *Comment penser la communication des connaissances? Du CD-ROM à l'Internet* (pp. 253-264). Paris : L'Harmattan.
- Dessus, P., Lemaire, B., et Baillé, J. (1997). Études expérimentales sur l'enseignement à distance ». *Sciences et Techniques Éducatives*, 4, 137-164.

- Foltz, P.W. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments and Computers*, 28, 197-202.
- Garrison, R. (2000). Theoretical challenges for distance education in the 21st century : A shift from structural to transactional issues. *International Review of Research on Open Distance Learning*, 1.
- Henri, F., et Ricciardi-Rigault, C. (1996). Collaborative distance learning and computer conferencing. In T.T. Liao (Ed.), *Advanced educational technology : Research issues and future potential* (pp. 45-76). Berlin : Springer Verlag.
- Kozma, R.B. (1991). Learning with media. *Review of Educational Research*, 61, 179-211.
- Landauer, T.K., et Dumais, S.T. (1997). A solution to Plato's problem : the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104, 211-240.
- Lemaire, B., et Dessus, P. (à paraître). « Modèles cognitifs issus de l'analyse de la sémantique latente ». *Cahiers Romains de Sciences Cognitives*.
- Lemaire, B., Marquet, P., et Baillé, J. (1997). Étude comparative du discours d'un enseignant et du comportement des enseignants en situation de cours traditionnel et d'enseignement à distance informatisé. *Carrefour de l'Éducation*, 3, 76-91.
- Lemaire, B., Dessus, P., & Baillé, J. (1998). The teacher discourse at a distance : Lexical, morphosyntactical and pragmatic aspects. *International Journal of Educational Telecommunications*, 4, 367-381.
- Lockee, B.B., Burton, J.K., et Cross, L.H. (1999). No comparison : Distance education finds a new use for « No Significant Difference ». *Educational Technology Research and Development*, 47, 33-42.
- Moore, M.G. (1993). Theory of transactional distance. In D. Keegan (éd.), *Theoretical principles of distance education* (pp. 22-38). New York : Routledge.
- Muller, C. (1992). *Principes et méthodes de statistique lexicale*. Paris : Champion.
- Ogata, H., et Yano, Y. (2000). Combining knowledge awareness and information filtering in an open-ended collaborative learning environment. *International Journal of Artificial Intelligence in Education*, 11, 33-46.
- Phipps, R., et Merisotis, J. (1999). *What's the difference? A review of contemporary research on the effectiveness of distance learning in higher education* (research report). Washington : Institute for Higher Education Policy.
- Reffay, C., et Chanier, T. (2001). CUMULI : Construction d'une mémoire du groupe dans l'interaction en FAD. *Sciences et Techniques Éducatives*, 8, 155-158.
- Russell, T.L. (2002). *The « no significant difference phenomenon »*. Site accédé le 28 octobre 2002 à <http://teleeducation.nb.ca/nosignificantdifference/>
- Saba, F., et Shearer, R.L. (1994). Verifying key theoretical concepts in a dynamic model of distance education. *American Journal of Distance Education*, 8, 36-59.
- Scardamalia, M., et Bereiter, C. (1994). Computer support for knowledge-building communities. *Journal of the Learning Sciences*, 3, 265-283.
- Teplovs, C., et van Aalst, J. (1998). Latent semantic analysis of CSILE/KF Databases ». *Proceedings of NCE Conference (NCE'98)*. Vancouver.

---

Philippe Dessus est maître de conférences à l'UFM de Grenoble, France, et chercheur au laboratoire des sciences de l'éducation de la même ville. Il y mène des recherches sur des applications de l'analyse de la sémantique latente visant à modéliser et aider les processus d'apprentissage et d'enseignement.