Notes de lecture

Rubrique préparée par Denis Maurel

Université de Tours, LIFAT (Laboratoire d'informatique fondamentale et appliquée)

Rotem DROR, Lotem PELED-COHEN, Segev SHLOMOV, Roi REICHART. Statistical Significance Testing for Natural Language Processing. *Morgan & Claypool publishers*. 2020. 98 pages. ISBN 978-1-68173-795-9.

Lu par **François LÉVY**

Université Sorbonne Paris-Nord / LIPN

J'avais, en choisissant ce livre, l'envie de comprendre ce que signifient vraiment les différents tests qui sont employés pour tirer des enseignements de ces tableaux de résultats sur lesquels nous appuyons une grande part de nos articles. C'était un peu trop optimiste : le livre est construit autour de trois articles parus dans TACL et à ACL et les explications ajoutées sont un peu succinctes pour aider le novice à comprendre en détail la méthodologie statistique. Il contient néanmoins une réflexion intéressante sur l'évaluation des expériences multiples qui jouent un grand rôle dans le travail actuel, et propose des méthodes pour cela. En cent pages, ce n'est pas si mal.

Le plan du texte découle de sa genèse. Trois chapitres présentent la problématique générale du test d'hypothèse, qui décide de la signification statistique. La suite est consacrée à la comparaison d'algorithmes. Le chapitre 4 fait une revue des principales tâches de traitement automatique des langues, des indicateurs servant à évaluer les résultats et des tests de signification utilisables selon le type d'indicateur pour comparer deux algorithmes à partir d'un unique corpus. La troisième partie traite également de la comparaison de deux algorithmes sur le même corpus, dans le cas particulier où ils sont tels qu'il faut un ensemble d'exécutions pour rendre compte de leur performance. Ce choix est motivé par les réseaux neuronaux profonds (RNP) dont les résultats varient avec l'initialisation. Une quatrième partie envisage la situation où la comparaison est faite sur plusieurs corpus : la problématique est très liée à la portabilité et à l'adaptation des logiciels à d'autres domaines ou d'autres langues.

La description générale du test d'hypothèse est indispensable à la compréhension. On a un ensemble d'éventualités d'une taille inaccessible à l'exploration, par exemple une série infinie de lancers d'une pièce de monnaie. Une variable aléatoire v évalue chaque éventualité ; disons -1 pour pile et 1 pour face. On a aussi un jeu de données v_{obs} observé sur une population et décrit par une statistique S_{obs}. On aimerait savoir

si ce jeu est significatif : si la moyenne S_{obs} des mille tirages observés est *0,074*, peut-on en déduire que la pièce est déséquilibrée vers face ? Le raisonnement ressemble au raisonnement par l'absurde. On considère deux hypothèses, H_1 affirmant que l'observation est significative, et H_0 supposant le contraire. La première étape est de modéliser S sous l'hypothèse H_0 par une distribution de probabilité – si la pièce est équilibrée, la distribution de la moyenne de mille tirages s'approxime par une loi normale. Ensuite, on considère la probabilité que S soit dans l'intervalle des valeurs au moins aussi significatives que l'exemple (ici $P(S \geq 0{,}074)$). Cette probabilité est la p-valeur de l'observation : elle mesure la probabilité qu'un résultat apparemment significatif soit seulement l'effet du hasard. Dernière étape, on choisit un seuil de confiance α : si la p-valeur est plus petite que $1\text{-}\alpha$, l'observation permet de rejeter H_0.

On voit bien, dans cette présentation abstraite, les conditions de rigueur derrière la technique. En premier, il faut justifier la distribution supposée de S dans les jeux de données sous l'hypothèse H_0, en particulier contrôler que les données vérifient les conditions de validité de la distribution (par exemple, l'indépendance). En second, la p-valeur ne dit pas tout de la qualité de la décision : c'est la probabilité de S_{obs} *connaissant* H_0 et on ne peut en déduire la probabilité de H_0 *connaissant* S_{obs}. Augmenter le seuil de confiance réduit le nombre de cas où l'on rejette H_0 à tort, mais augmente ceux où l'on ne détecte pas que H_0 devrait être rejeté ; minimiser l'erreur est plus compliqué qu'il n'y paraît. D'un point de vue opérationnel, on appelle tests paramétriques ceux qui utilisent la distribution de S; quand celle-ci est inconnue, on peut se rabattre sur des tests dits non paramétriques : soit on utilise un indicateur dérivé moins précis, soit on évalue empiriquement la distribution. Quelques exemples de tests constituent l'essentiel du chapitre 3 ; ils sont trop succinctement expliqués pour apprendre grand-chose à qui ne les connaît pas déjà, ce qui fait de ce chapitre pour l'essentiel une indirection vers d'autres sources.

Le premier thème abordé ensuite est donc une revue des tests de signification servant à comparer deux algorithmes dans les différentes tâches de traitement automatique des langues. Les éventualités sont implicites ; on peut pour certaines tâches identifier dans ce rôle l'ensemble des phrases ou l'ensemble de textes possibles. Pour un jeu de données, la statistique est la différence δ des indicateurs mesurant la performance de chacun des algorithmes à comparer (par exemple, leur précision) et, sans perte de généralité, l'hypothèse H_1 est que δ est positif. Les valeurs au moins aussi significatives que l'exemple vérifient alors $\delta \geq \delta_{obs}$. L'essentiel est un tableau d'une vingtaine d'indicateurs utilisés dans de nombreuses tâches de TAL, avec les tests paramétriques (si possible) et non paramétriques qui conviennent pour cet indicateur. À noter les indications de rigueur qui figurent en renvoi, par exemple, que les tests paramétriques supposent une distribution normale de l'indicateur. Et la dernière notation du chapitre vaut la peine d'être signalée : quand la différence de performance entre les algorithmes est faible, on peut changer la réponse du test en augmentant la taille du jeu de données observé ! Cela s'appelle du « *p-hacking* ».

Le thème de la comparaison des réseaux neuronaux profonds est plus récent. Le constat de départ est simple : même à ensemble d'apprentissage et de développement constants, une seule exécution du réseau ne suffit pas à déterminer un indice de performance fiable sur l'ensemble de test. Ceci est dû à ce que les auteurs appellent le non-déterminisme des RNP : le grand nombre de paramètres de structure, la variabilité de leurs méthodes d'optimisation, les choix aléatoires (par exemple, d'initialisation des poids ou d'ordre de présentation des exemples) peuvent, toutes choses égales par ailleurs, produire des variations de performance importantes. Il est donc nécessaire pour comparer deux algorithmes de les caractériser chacun par un ensemble d'exécutions. En passant, la définition implicite des éventualités a changé par rapport au premier thème : ce sont les exécutions possibles sur un seul jeu de données. Les auteurs critiquent les méthodes de comparaison simples. Ils proposent à la place un affaiblissement de l'ordre stochastique naturel du premier ordre[1], dont ils estiment le critère trop strict. Ils mesurent pour cela à quelle distance εW_2 (entre *0* et *1*) une des variables peut dominer l'autre. On peut alors calculer le meilleur majorant de εW_2 pour les ensembles d'exécutions et un seuil de confiance donné. L'analyse empirique comporte des expériences avec un bruit contrôlé et cinq cent dix comparaisons extraites de cinq tâches différentes. Elle incite à poursuivre cette piste.

Le dernier thème concerne la comparaison de deux algorithmes sur plusieurs jeux de données. Une remarque liminaire pose le problème : si dix jeux rejettent chacun H_0 avec un seuil de confiance de *0,95*, le seuil de confiance de l'appréciation globale « A est dans les dix cas meilleur que B » est un peu en dessous de *0,6*. Ici, les éventualités sont donc les types de jeux de données (type de texte, langue, etc.). Les auteurs utilisent des méthodes développées pour le traitement d'images médicales. S'il y a N jeux de données, on considère une famille d'hypothèses nulles ${H_0}^{u/N}$ considérant que moins de u hypothèses nulles parmi celles attachées à chaque jeu de données individuel sont fausses. Comme on connaît les p-valeurs associées aux hypothèses nulles individuelles, les auteurs proposent deux formules pour calculer celles associées aux ${H_0}^{u/N}$ selon que les jeux de données sont garantis d'être indépendants ou pas. Il est alors possible de calculer le nombre d'hypothèses nulles que l'on peut rejeter avec le seuil de confiance global α. Dans un deuxième temps, une procédure fournit une liste de jeux de données pour lesquels on peut rejeter l'hypothèse nulle avec le même seuil de confiance global (dans le cas d'indépendance, moins que ce que l'on avait calculé précédemment). Ici aussi, une série d'expériences complète l'argumentation. Les trois solutions comparées donnent des résultats différents, mais les critères d'indépendance qui permettent d'en choisir un sont très elliptiques.

En conclusion, ce livre traite d'un problème important pour le traitement automatique des langues : quand on compare des algorithmes, comment interpréter rigoureusement les résultats obtenus sur un corpus particulier. Et il propose des

1 Pour les variables aléatoires réelles X et Y, $X \geq_{1a} Y$ si, et seulement si, pour tout a, $P(X > a) \geq P(Y > a)$.

solutions pour des configurations dont l'importance est récente, les algorithmes non déterministes et les évaluations multicorpus. Il signale aussi les points encore obscurs dans l'utilisation du test d'hypothèses par le traitement automatique des langues, au premier rang desquels la difficulté à décider de l'indépendance des données qui joue pourtant un rôle crucial. Il me semble qu'une explicitation plus détaillée de l'ensemble des éventualités sur lequel on raisonne aiderait à clarifier la question.

Shashi NARAYAN, Claire GARDENT. Deep Learning Approaches to Text Production. *Morgan & Claypool publishers*. 2020. 176 pages. ISBN 978-1-68173-758-4.

Lu par **Caio Filippo CORRO**

Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique (LISN)

Cet ouvrage propose une vue d'ensemble sur la recherche en génération automatique de textes fondée sur des réseaux de neurones. Il est découpé en trois parties couvrant (1) les modèles fondamentaux, (2) les méthodes neuronales avancées et enfin (3) les jeux de données disponibles. L'ouvrage est agréable à lire et contient de nombreuses illustrations et exemples. Il ne nécessite pas de prérequis sur le domaine et conviendra donc parfaitement à un stagiaire de master ou un doctorant voulant découvrir les approches modernes de génération automatique de textes.

Génération automatique de textes et réseaux de neurones

Le domaine de la recherche sur la génération automatique de textes s'intéresse à un ensemble très hétérogène de tâches qui ont toutes en commun de produire des énoncés en langage naturel, mais qui diffèrent à la fois sur le type de source d'information que l'on utilise en entrée du modèle et les objectifs attendus pour la sortie. L'entrée peut être une phrase, un document complet, une représentation sémantique ou encore des données brutes. Les sorties attendues peuvent être soit dans la même langue que l'entrée, soit dans une langue différente (traduction automatique). Certaines tâches portent beaucoup d'importance sur les propriétés du texte généré. Par exemple, pour la simplification de textes, l'objectif est de réécrire une ou plusieurs phrases afin de les rendre plus accessibles. Bien que l'idée du « modèle unique » qui s'appliquerait à tous les cas de figure soit séduisante, en pratique cela ne fonctionne pas. L'ouvrage propose un panorama des architectures neuronales et des méthodes d'entraînement permettant de prendre en compte ces différentes spécificités.

Contenu de l'ouvrage

La première partie de l'ouvrage comprend deux chapitres introduisant les grandes généralités sur la génération automatique de textes. Le chapitre 2 est un résumé très bref (une dizaine de pages) sur les méthodes non neuronales. Le

chapitre 3 détaille l'architecture encodeur décodeur standard de l'approche neuronale :

- l'encodeur transforme l'entrée, par exemple une phrase ou une représentation sémantique, en un vecteur de taille fixe aussi appelé plongement ;
- le décodeur utilise ce vecteur pour conditionner la génération d'une phrase avec un modèle autorégressif.

Ce chapitre décrit brièvement, mais efficacement, le fonctionnement des réseaux de neurones récurrents, les plongements lexicaux et les méthodes d'entraînement. Cela peut tout à fait faire office d'introduction à l'apprentissage profond pour un apprenti chercheur dans ce domaine. Le chapitre se termine par une brève comparaison avec les méthodes présentées dans le chapitre 2.

La seconde partie de l'ouvrage, certainement la plus importante, présente les différentes modifications pouvant être apportées à l'architecture encodeur décodeur. Le chapitre 4 s'intéresse aux mécanismes qui permettent d'améliorer les interactions entre l'encodeur et le décodeur : attention, copie et contraintes de couverture. Ces trois mécanismes sont très liés et sont fondés sur l'idée d'introduire un lien direct entre les mots de l'entrée et la distribution sur le vocabulaire de sortie à chaque position. L'attention permet de créer « un raccourci » entre l'entrée et la sortie, ce qui est particulièrement important, car le plongement produit par l'encodeur est de taille fixe, ce qui induit des difficultés pour les entrées longues et complexes. La copie permet de faire un copier-coller des termes de l'entrée, par exemple pour générer des noms propres qui ne sont pas vus lors de l'entraînement et donc sont non présents dans le vocabulaire du décodeur. Enfin, la couverture permet d'empêcher le décodeur de produire des contenus redondants ou d'oublier des parties importantes de l'entrée. Le chapitre est suffisamment détaillé pour permettre à celui qui le souhaite de travailler directement sur une implémentation de ces trois mécanismes.

Le chapitre 5 présente les différents types d'architectures neuronales pouvant être utilisés au niveau de l'encodeur pour s'adapter à la structure de l'entrée. Bien que de nombreux travaux dans la littérature se focalisent uniquement sur les réseaux de neurones récurrents, peu importe la structure de l'entrée, ces derniers rencontrent des difficultés en pratique. Dans le cas de textes longs, une approche naïve consiste à simplement concaténer l'ensemble des phrases en une unique séquence. Les réseaux de neurones récurrents se heurtent alors au problème de la prise en compte des dépendances à longue distance. En effet, ceux-ci doivent encoder un historique arbitrairement long dans une représentation de taille fixe. L'ouvrage présente des approches de modélisation fondées sur des architectures neuronales hiérarchiques où la structure du document (découpage en phrases, en paragraphes) est prise en compte pour contrer ce problème. Dans le cadre d'entrées qui ne sont pas des énoncés en langage naturel, comme dans le cas des représentations sémantiques qui ont une forme de graphe, l'approche dite de linéarisation consiste à représenter ce graphe sous forme d'une séquence. Ceci pose des problèmes de pertes d'informations (certaines relations ne sont pas représentées dans la linéarisation) et de localité (des nœuds proches dans le graphe peuvent se retrouver éloignés dans la

linéarisation). L'ouvrage décrit les approches fondées sur des réseaux de convolution de graphes pour encoder directement les représentations sémantiques sans linéarisation artificielle.

Enfin, le chapitre 6 s'intéresse à la prise en compte des objectifs attendus sur le texte produit. Par exemple, dans le cas de la génération de résumés, il est important de considérer des systèmes en deux étapes qui commencent par repérer les contenus saillants avant de s'attaquer à la génération proprement dite. Ce chapitre s'intéresse aussi à l'utilisation des métriques d'évaluation comme objectif à optimiser lors de l'apprentissage. Ces métriques sont pour la plupart non différentiables. L'ouvrage se concentre sur les approches d'apprentissage par renforcement qui permettent de contourner ce problème en utilisant directement le score donné par la métrique comme signal d'apprentissage. Les autres méthodes comme la construction de substituts différentiables ne sont pas abordées.

La troisième et dernière partie de l'ouvrage présente en un unique chapitre les jeux de données disponibles pour les différentes tâches de génération automatique de textes.

Conclusion

Cet ouvrage propose une vue d'ensemble sur la recherche en génération automatique de textes fondée sur des réseaux de neurones. Un soin particulier a été apporté au niveau de la pédagogie : très peu de prérequis sont nécessaires, de nombreuses figures permettent de comprendre rapidement les concepts décrits dans le texte (le chapitre 3 est d'une exemplarité rare dans ce domaine !) et les équations sont clairement détaillées. Cet ouvrage pourrait être un très bon support pour un cours niveau master sur le sujet.

Cependant, il décevra probablement les chercheurs aguerris. Les différents thèmes ne sont couverts que de façon sommaire. Le livre ne contient ni analyse théorique ni analyse expérimentale. Par exemple, la section sur l'apprentissage par renforcement contient une explication claire qui permet de bien comprendre pourquoi cette approche n'est pas dépendante directement du gradient des métriques d'évaluation pour entraîner un modèle. Cependant, la forte variance de cet estimateur n'est pas expliquée formellement et n'est que brièvement discutée. Aucune analyse des méthodes de réduction de la variance et de leur impact sur le temps d'entraînement et les résultats expérimentaux n'est proposée.

Jacques MOESCHLER. Pourquoi le langage ? Des Inuits à Google. *Armand Colin*. 2020. 286 pages. ISBN 978-2-200-62855-0.

Lu par **Alain MILLE**

Université Lyon 1 / LIRIS UMR CNRS 5205

L'auteur s'appuie sur les théories linguistiques, cognitives et pragmatiques pour s'interroger sur ce que l'on sait et ce que l'on ne sait pas sur le langage. Il ne répond pas à la question du

titre « Pourquoi le langage ? », mais brosse un tableau de toutes les questions qu'il faut étudier pour répondre à cette simple interrogation. L'ouvrage a une facette pédagogique, avec beaucoup d'exemples et de contre-exemples pour illustrer la complexité de la tâche. Il manipule souvent la démonstration par l'absurde pour démontrer que les théories standard ne suffisent pas et il introduit progressivement la théorie pragmatique qu'il défend comme une approche permettant de comprendre le langage en contexte. Certaines approches, essentiellement les approches en neurolinguistique, ne sont pas abordées dans l'ouvrage qui donne toutefois une bonne vision de l'état des connaissances actuelles pour répondre à la question « Pourquoi le langage ? ».

Notice d'ouvrage

En avant-propos, l'auteur livre ses intentions : 1) parce que le langage est important et certainement la propriété unique qui définit notre espèce, 2) parce que le langage est généralement considéré comme allant de soi, et que pourtant si les connaissances sont maintenant importantes elles sont encore très lacunaires. Il y a donc d'une part un discours de pédagogie pour montrer l'importance du langage et de ce qu'il révèle de la nature humaine et, d'autre part, une sorte d'état de l'art des connaissances en la matière, en assumant le fait de ne pas aborder les travaux en neurosciences cognitives ou en linguistique informatique.

L'introduction reprend ces objectifs en précisant le public cible : toute personne s'intéressant au langage qu'il soit scientifique ou non. L'auteur annonce qu'il présentera surtout ce qu'il connaît bien et fera un focus sur ses propres travaux en pragmatique. Il donne plusieurs raisons qui font qu'il est difficile de traiter du langage dans un ouvrage unique : 1) chaque locuteur se sent expert de sa propre langue, 2) une approche scientifique du langage est considérée avec scepticisme par certaines disciplines, 3) l'idée que la culture et la langue française seraient supérieures aux autres, d'autant que les autres doivent penser la même chose de leur propre langue… Une bonne partie du reste de l'introduction est consacrée à la démonstration de la valeur et de l'importance de la linguistique comme discipline scientifique. Il pose que l'hypothèse relativiste (pas d'invariants linguistiques universels) l'emporte sur l'hypothèse que toutes les langues seraient des variations d'un même patron, une grammaire universelle. Il annonce deux thèmes majeurs qu'il abordera de manière plus approfondie : la différence entre étude du langage et étude de la communication, et la valeur d'une approche pragmatique qu'il défend en tant que chercheur engagé dans cette voie.

Partie I

La partie I traite du lien entre langage et communication.

Dans le chapitre 1, l'auteur énonce ce qui constitue à son avis huit idées fausses sur le langage : 1) les langues non écrites ne seraient pas de « vraies » langues, 2) il y aurait des langues plus « importantes » que d'autres, 3) le français serait une langue logique, claire et belle, 4) les langues souffriraient de l'influence d'autres langues, 5) il faudrait protéger le français de l'influence des autres langues, 6) les enfants apprendraient leur langue maternelle par imitation des paroles de leurs parents, 7) seuls les mots du dictionnaire appartiendraient à la langue et 8) le

linguiste s'intéresserait à l'origine des mots. Il conclut par les propositions alternatives consistant à nier les idées fausses.

Remarque sur ce chapitre : la réfutation de l'idée fausse 6 est l'occasion d'introduire un principe qui sera repris largement dans le reste de l'ouvrage : « *les enfants n'apprennent pas seulement par la langue maternelle par imitation de leurs parents, mais naturellement parce qu'ils sont programmés pour cela* ». On peut s'interroger sur le terme de *programmation.*

Le chapitre 2 se présente avec une volonté démonstrative de l'affirmation que le langage n'est pas la communication et la communication n'est pas le langage. Après avoir fait une présentation pédagogique de la communication, l'auteur introduit deux modèles : un modèle *codique* basé sur les codes linguistiques et les codes sociaux et un modèle *inférentiel* où les informations non formulées sont déduites du contexte. De la même façon la notion de langage est présentée laconiquement : « *Un langage est constitué d'une phonologie, d'une sémantique et d'une syntaxe* ». Deux fonctions du langage sont affirmées : communication et cognition. Si l'aspect communication a été déjà anticipé, l'aspect cognition est démontré par la propriété que le langage a d'enchâsser une structure, quelle qu'elle soit, dans une structure du même type. La communication ne serait qu'un effet de bord d'une émergence du langage comme support du développement de la cognition : le langage comme support de la pensée. Cette externalisation de la pensée permet la réflexion et la récursivité qui expliqueraient le lien entre la pensée, le langage et le raisonnement.

Le chapitre 3 défend l'idée que ce n'est pas la structure du langage qui en définit l'usage. C'est ici le thème de prédilection de l'auteur : la pragmatique comme étude de l'usage du langage dans la communication. L'apprentissage de la langue ne se fait pas à partir des règles du langage. La conversation obéit à des règles qui ne sont pas liées spécifiquement au langage. À l'instar de Noam Chomsky, la distinction est faite entre compétence et performance du langage. Une distinction est ainsi introduite entre langue interne (li) et langue externe (le). L'auteur interroge l'agenda de la recherche sur la question en se demandant s'il faut être compétent (connaître finement la langue) pour être performant (utiliser finement la langue). À la suite de Chomsky, il est admis que le système cognitif (*Faculty of Language in the Narrow sense*, FLN) interagit avec une faculté (*Faculty of Languange in the Broad sense*, FLB) s'appuyant sur deux systèmes « externes » : le système articulatoire perceptuel et le système conceptuel intentionnel. FLB serait une évolution d'une faculté partagée avec beaucoup d'animaux tandis que FLN serait une faculté apparue plus récemment et spécifique de l'humain. Il y aurait donc eu un protolangage (langage sans syntaxe) avant le langage moderne. L'auteur décrit comment la compréhension de la communication verbale a été modifiée par l'hypothèse d'un principe de coopération et de maximes de conversation, qui sont liées à la cognition. De nombreux exemples émaillent cette partie pédagogique de l'ouvrage. Cette section permet de montrer qu'il faut ajouter un principe de pertinence pour expliquer la communication verbale. Cette théorie de la pertinence est traitée largement, en particulier pour le cas de la communication implicite.

Partie II

La partie II s'intéresse à la partie sociale du langage.

Dans le chapitre 4, l'auteur s'interroge à nouveau sur ce qui serait caractéristique de l'humain, le langage n'étant pas seul candidat. C'est dans ce chapitre que l'auteur développe la théorie qu'il va ensuite affirmer à partir de l'hypothèse SAPIR-WHORF qui articule le langage avec la culture *via* le vocabulaire qui se stabilise. Cette relation met en évidence qu'il existe des mots intraduisibles car en relation avec des éléments culturels disjoints. Dans l'approche pragmatique, la valeur d'un mot ne prend son sens que dans son usage en contexte. Les variations linguistiques sont traitées avec l'exemple du français et du « vernaculaire noir-américain ». Le chapitre se termine par des mises en évidence des formes de politesse, de face et de figuration qui sont autant de formes étudiées dans la communication verbale.

Dans le chapitre 5, l'auteur commence par démontrer qu'il n'y a pas de règles spécifiques au discours par un raisonnement par l'absurde mobilisant des contre-exemples. Il conclut qu'aucune des règles souvent admises n'est générale ni constructive de la notion de discours. Cette démonstration faite, l'auteur propose sa théorie d'une pragmatique du discours. La cohérence est un jugement de l'observateur du discours, pas une propriété intrinsèque du discours : c'est la conséquence du processus de compréhension. Dans le cadre de la théorie de la pertinence, la compréhension est liée à deux effets sur les croyances : effets non propositionnels et effets propositionnels. Les effets non propositionnels sont liés aux émotions et à l'adhésion aux éléments du discours. Les effets propositionnels, au contraire, sont liés à l'argumentation et à la persuasion. Le discours n'est pas une unité linguistique, mais une unité pragmatique.

Dans le chapitre 6, une introduction s'interroge sur la fonction des métonymies, des métaphores, des images mentales dans le langage. L'auteur réfute les fonctions classiques qu'on leur attribue, ce qui lui permet d'introduire le modèle explicatif de la théorie de la pertinence. Cette théorie fournirait des clés nouvelles pour comprendre ces usages. C'est le contexte du discours émis ou observé qui fournit le sens, pas la valeur littérale des expressions utilisées, même si elles sont ancrées dans la culture partagée. La structure de récit est également remise en question comme n'étant pas une structure de communication, mais caractérisée par un ordre des événements dans un style indirect libre encore une fois lié essentiellement à une stratégie linguistique pour prêter des paroles ou des pensées à une troisième personne. L'auteur termine le chapitre en introduisant la notion de super-pragmatique, au-delà de la pragmatique déjà traitée dans les chapitres précédents.

Dans le chapitre 7, en exploitant l'idée de la super-linguistique, l'auteur propose l'idée de super-pragmatique pour utiliser la méthode de la pragmatique pour aller au-delà du domaine d'investigation de la pragmatique. C'est ainsi qu'il se propose d'exploiter deux concepts centraux de la pragmatique, présupposition et implicature, pour comprendre des enjeux de la société et ses crises. La compréhension d'un énoncé mobilise plusieurs mécanismes : implicature conversationnelle, implicature conventionnelle, présupposition et implications logiques. L'auteur propose d'ajouter la notion « d'explicature » qui peut expliciter directement la proposition exprimée

dans l'énoncé, mais aussi, à un niveau supérieur, la valeur d'acte de langage de l'énoncé. Ces mécanismes varient en termes d'accessibilité et de force. Les rôles des présuppositions puis des implicatures sont interrogés. Les présuppositions doivent être partagées et les implicatures peuvent être annulées par des implications qui s'imposent. Pour illustrer l'intérêt de la super-pragmatique, l'auteur s'intéresse à la posture de décryptage adoptée par les journalistes qui formulent des implications sur les informations qu'ils traitent. Le cas d'usage de « *Je suis Charlie* » est utilisé pour montrer que si « *Je suis Charlie* » alors cela implique un point de vue relativement partagé en solidarité avec les victimes de l'attentat, le « *Je ne suis pas Charlie* » n'implique par un point de vue partagé avec les assassins pour autant. L'auteur conclut le chapitre par cette phrase : « *Ce qui manque crucialement aujourd'hui, c'est une réflexion approfondie sur le langage et les implications de son usage par la presse en générale et les journalistes en particulier* ».

En conclusion, l'auteur distingue ce que nous savons et ce que nous ne savons pas encore sur le langage. *Ce que nous savons (imparfaitement) :* de la syntaxe à la pragmatique. La linguistique est complexe, car une part fondamentale repose sur sa nature orale, dynamique et sur le contexte échappant à l'énoncé. *Ce que nous ne savons pas encore :* émotions, origine du langage, traduction automatique, et communication homme-machine.