

LA PROGRAMMATION DU SENS

Jean-Pierre BALPE

1) Un exemple

Sur les murs d'une ville, une affiche : elle porte sur un fond totalement noir, en grosses lettres grises l'inscription suivante : « *Le problème avec le dernier verre c'est que c'est parfois le dernier* ». Tout locuteur français comprend cette affiche. Elle a pour lui un sens évident. Et pourtant, un examen plus attentif en révèle toutes les ambiguïtés : prise au pied de la lettre, en ne tenant compte que de ce que disent les mots qu'elle porte, elle énonce une tautologie sous forme de pléonasme : « *un dernier verre est un dernier verre* », comme « *H₂O = H₂O* »... Or tout lecteur de cette affiche sait que le sens exprimé n'est pas celui-là car ce sens-là ne l'intéresserait pas. En fait, ce qu'il comprend est plus complexe : il perçoit immédiatement le *jeu sur les mots* qui fonde le sens et repose sur la *métonymie* : ce qui est en jeu, n'est pas le verre, mais la consommation éventuelle de son contenant ; d'autre part « *un dernier verre n'est pas un dernier verre* ». Il y a plusieurs « derniers verres » possibles, autant que d'échelles temporelles : un verre peut-être le dernier d'un moment donné, or la liste de ces moments possibles est aussi infinie que les possibilités de fractionnements temporels. C'est ce que, *sans le dire*, affirme le slogan. Et c'est parce que *cela va sans dire* que ce slogan pose problème car le sens qu'il porte n'est pas dans les mots, mais dans un extérieur à ces mots que le lecteur convoque. Ce que le lecteur comprend est à la fois un avertissement — trop boire est dangereux : un *dernier verre* implique une série indéfinie de verres vidés dans une courte période donnée —, un conseil — ne buvez pas trop car vous risquez un accident, une constatation : toute vie a une fin — et une position

philosophique — la vie vaut la peine d'être vécue le plus longtemps possible. La chaîne sémantique qui s'établit est donc la suivante : si vous buvez trop dans une période courte de temps et que vous preniez votre voiture, alors, parce que vous serez ivre, donc moins lucide, vous risquez un accident qui peut vous coûter la vie et ce serait trop bête de mourir pour cette raison-là. Ce qui se cache *sous les mots*, est à la fois une connaissance et une philosophie du monde qui ne pourraient s'exprimer sans les mots, avec d'autres mots, mais pour laquelle ces mots-là ne sont qu'un *pré-texte*.

Il serait facile de multiplier de tels exemples et de trouver pour chacun une explication *ad-hoc* : le sens de tout texte est bien plus en-dehors du texte que dans le texte lui-même. Pour autant, sans texte, sans tissage de liens, il ne peut y avoir de sens.

2) Qu'est-ce que le sens ?

Le sens est un tissage de relations. Lorsque du sens apparaît, c'est-à-dire lorsqu'un locuteur est capable, au-delà de la signification d'un texte, d'en tirer des conséquences pragmatiques, ce qui se construit, c'est une organisation de relations entre objets du texte et du monde : sans connaissance des pratiques consommatrices en terme de boissons, sans connaissance des divers types de boissons, sans connaissance des effets sur l'homme des boissons alcoolisées, sans connaissance des modes de déplacements humains, sans connaissance des véhicules automobiles, etc... le slogan cité est asémantique. Le sens n'est pas un « donné en soi », il n'existe pas **DANS** la langue mais **À TRAVERS** elle qui n'est qu'un instrument de médiation. C'est pour cette raison précise que toute langue est apte à porter n'importe quel sens, y compris des sens contradictoires. Tout établissement de sens repose sur un tissage de relations effectives dans un ensemble de relations possibles : le sens est une mise en contextes, mise en réseau d'informations, de significations et de connaissances. Ce qui entre en jeu est un enchevêtrement complexe de systèmes de mises en relations. Pour aller vite, un ensemble de structures emboîtées de con-textualisation qui peut être décrit ainsi... (voir tableau page suivante)

texte	relations entre les signes réellement présents dans un message particulier donné
phénotexte	relations entre les signes du message et tous les signes de tous les messages du même auteur
intertexte	relations entre les signes du message et l'ensemble des messages du même type
péritexte	relations entre les signes du message et son entour pragmatique immédiat
métatexte	relations entre les signes du message et l'ensemble pragmatique des mondes

Le sens se calcule toujours entre le prévisible (les choses sont déjà été dites, connues et immédiatement acceptées) et l'aléatoire (les choses n'ont pas encore été dites et sont de l'ordre d'un possible ouvert) : une partie en est *toute faite*, alors qu'une autre se construit. Ce qui implique une dissymétrie : comme le montrent à l'envie toutes créations fictionnelles ainsi que les usages quotidiens de chacun produire du sens et interpréter du sens ne sont pas deux mécanismes inversés faisant appel *en miroir* aux mêmes opérativités.

3) Programmer du sens

L'informatique ne traitant que des formes descriptibles, tout le problème de la programmation du sens est là, à la fois dans l'asymétrie fonctionnelle et dans les possibilités d'établissements de divers niveaux de relations.

3.1) Le traitement du texte

Si *programmer du sens*, consiste à produire ou analyser du texte, c'est-à-dire une forme, alors cela ne pose pas de problème majeur. En effet, analyser du texte et produire du texte consistent à ne reconnaître ou construire que des formes. Bien entendu, il y a un certain niveau de complexité bien connu des spécialistes du traitement automatique du langage. En voici quelques exemples...

- les formes et leurs relations internes sont propres à chaque langue et arbitraires : *dog* et *chien* sont et ne sont pas le même mot ; le français connaît des relations d'accord qu'ignore l'anglais, *pretty* ne change pas devant *dog* ou *dogs*, alors que *beau*, reste *beau* devant *chien* et devient *beaux* devant *chiens*,

etc... Mais la totalité de ces formes est descriptible — c'est le rôle des *dictionnaires* — et la totalité, ou presque de leurs relations l'est également — c'est le rôle des *grammaires*.

- dans certaines langues comme le français, les formes sont polysémiques, sous une même forme peuvent se cacher des *entrées* différentes : *livre* par exemple diffère suivant son contexte immédiat, *il livre, le livre, la livre...* Cette difficulté oblige à concevoir des analyseurs disant soit la nature syntaxique de la forme, soit conservant les multiples sens possibles, mais elle n'est pas rédhibitoire. Elle implique simplement que les dictionnaires utilisés comportent des informations intralinguistiques comme la nature syntaxique et une certaine forme de synonymie.

- il n'y a pas homogénéité entre formes (entrée d'un dictionnaire) et mots (définis, en français par exemple, par un ensemble de lettres entre deux signes de ponctuation). La plupart des formes sont *composées* et cette composition peut revêtir des aspects divers : radical-terminaison dans la plupart des cas — formes = forme + s — mais aussi adjonction d'une séquence indéterminée de formes : un *chien assis* n'est pas plus un chien qui s'est assis qu'un *hot dog* n'est un chien chaud, et aucune règle ne permet de savoir si la forme totale est composée de deux ou x formes élémentaire comme dans *garde-boue, corps de garde* ou *garde du corps...* Cette difficulté complique l'analyse d'une part parce qu'elle multiplie les synonymies, d'autre part parce qu'elle oblige à accroître, de façon considérable les entrées, mais elle n'est pas de l'ordre de l'impossible... La difficulté principale réside dans la dynamicité du système linguistique qui, chaque jour, produit de nouvelles entrées.

- l'ensemble des relations possibles dans un texte entre formes et l'incidence de ces relations sur les formes — les *règles* de syntaxe —, s'il n'est pas totalement ouvert, comme le montrent bien les faiblesses des analyseurs syntaxiques, est complexe, parfois mal fixé (exceptions, tolérances, etc...) mais descriptible pour plus de 90% des textes.

Ces difficultés concernent essentiellement l'analyse. La production du texte peut en effet être réalisée à partir d'un sous-ensemble restreint de dictionnaires et de syntaxes sans que son lecteur n'y trouve à redire. Le lecteur

est en effet habitué à produire des textes à partir de sous-ensembles, aucun locuteur d'une langue n'en connaissant ni tous les mots ni toutes les possibilités syntaxiques. Cependant il faut souligner que ces niveaux de traitement ne produisent pas du sens : ils produisent une description formelle de la langue ; en ce qui concerne la production, ils produisent une séquence acceptable en langue sans aucune idée de ce qu'elle peut bien vouloir signifier. Aussi un générateur maîtrisant les règles symétriques : connaissance des termes, des groupes de termes et de leurs relations d'accord, peut écrire : « *le couple avait recourci la vengeance* » ou « *le cheval avait entêté le synonyme* ». Il est à remarquer que, dans ce cas, il est très génératif puisque, à partir d'un dictionnaire donné et de quelques règles, il produit un nombre infini de séquences.

3.2) Le traitement du métatexte

Ceci montre bien les manques : pour qu'un analyseur aille au-delà d'une simple description formelle, pour qu'un générateur produise des séquences susceptibles d'avoir du sens, il faut dépasser le texte et envisager l'ensemble des autres niveaux de contextes et notamment considérer les relations formes-monde donc considérer le monde comme un ensemble d'objets en relations ou objets ET relations peuvent être formellement définis.

Deux exemples encore : Représenter un chien en terme de métatexte consiste à dire que...

- le chien est un animal généralement domestique
- le chien est représenté par des variétés de formes appelées races
- le chien a quatre pattes
- le chien est omnivore, mais plutôt carnivore
- le chien est un animal diurne
- etc...

Évidemment une telle description n'est opératoire que si chacun des termes qui définissent les liens et les nœuds de liens font partie du réseau. Ainsi dire que le chien *est utilisé pour* la chasse suppose que le terme chasse soit défini comme (Petit Larousse) : « *action de poursuivre, de prendre et de tuer le gibier* ». Tous termes qui doivent figurer à leur tour dans le réseau.

Représenter *l'aboïement du chien* en terme de formes est possible. Il suffit d'établir qu'il existe une relation entre une classe d'animaux et une production sonore audible par l'homme, relation générique qui permet aussi de traiter « la pie jacasse », « l'éléphant barrit » ou « la vache meugle » et de déterminer que dans le cas de l'application de cette relation à la classe des animaux désignés par la forme *chien*.

Il y a bien entendu autant de cas d'applications que de sous-classes dans la catégorie *des animaux à cris* que de cas intermédiaires, par exemple dire que l'oie *cacarde*, oblige à distinguer au moins trois classes d'oiseaux : ceux qui ne chantent pas, ceux qui chantent et ceux — dont l'oie — qui ont un cri spécifique. Ces ensemble de descriptions relationnelles constituent autant de *représentations de connaissances* sur le monde. La puissance d'un programme sémantique dépend strictement de l'ensemble de ces représentations de connaissances. Et là est bien la difficulté qui n'est pas de l'ordre du théorique mais du pragmatique : la possibilité de description se heurte à l'infini du réel et à sa mobilité. Si le monde était fini, fermé, une telle approche, même si elle prenait du temps serait envisageable — et elle l'est d'ailleurs tout à fait dans le cas de micromondes spécialisés et bien définis — mais elle ne l'est pas dans un monde ouvert et dynamique où les relations entre objets du monde ne cessent de se reconfigurer. L'apparition récente, par exemple, des pitbull et autres rottweiler ainsi que des nouvelles relations maîtres-chiens et chiens-public qui y sont liées oblige à remodeler une part importante des relations. Les objets du monde et les relations que les objets entretiennent dans le monde sont dans une reconfiguration permanente. Pour obtenir une programmation du sens aussi efficace que celle réalisée par le cerveau humain, il faut un programme qui ait des caractéristiques humaines, c'est-à-dire qui, captant sans cesse des informations, soit capable de reconfigurer sans cesse ses représentations. Dans ce cas, à moins d'être une *intelligence collective*, c'est-à-dire de ne négliger aucune information émise où que ce soit et n'importe quand, cette programmation aura également les défauts humains de la non-exhaustivité et de la non-homogénéité : elle ne permettra pas une maîtrise sémantique de tout sur tout et comportera des zones spécialisées.

4) Produire du sens

Il n'est pas ici possible d'examiner l'ensemble des problèmes liés à la programmation du sens. La suite de l'exposé sera donc centré sur un aspect particulier du problème, celui de la génération automatique.

Un générateur automatique est un programme particulier qui, à partir d'algorithmes et de données, peut rédiger le texte suivant : « *Crépuscule bleu, tombée de nuit, ciel noir sombre : le vent emplit l'espace de son poids horrible ! Alors que le jour tombe, sous les vociférations du peuple, sous des milliers de regards agressifs, les nuages se déchirent ! Trois charrettes débouchent sur la place. Un nuage se met à couvrir le soleil, lent, large et gris; gris, le cercle du soleil tourne. Les Tuileries et les Champs Elysée regardent, la nuit tombe comme un couperet, des barbotteuses font de l'œil aux hommes qui passent ! Trois chariots peints de rouge débouchent sur la place. Sur une des charrettes une condamnée harangue le peuple, un autre chante une chanson, un autre encore menace. Il fait noir ; une jeune fille, immobile, descend du chariot, monte lentement les escaliers de l'échafaud, s'approche du bourreau et de ses aides, grand ciel rouge, nuages ; des voix s'élèvent : « sale putain, boucanense ! » Les nuages obscurs couvrent la ville d'un voile de deuil, les aides du bourreau s'emparent de la condamnée ! Le soleil lance dans les nuages de grands jets de sang. À son tour le bourreau rouge s'avance, approche de la guillotina, libère le couperet. Le ciel traîne. La tête coupée roule sur l'échafaud ! Le bourreau se tourne vers le peuple comme en quête d'applaudissements ! Le sang coule en abondance sur le pavé - le soleil lance de grands jets de sang ; le corps est jeté dans une carriole... Des cris traversent la foule — la chaleur est terrible ! L'espace sombre du ciel lourd enferme l'âme — sur le chariot plusieurs condamnés pleurent... »*

Si l'on s'en tient à ce qui vient d'être dit, produire de tels textes — issus du roman TRAJECTOIRES installé à l'adresse trajectoires.com à partir de décembre 2000 — est une gageure. En effet, le programme génératif semble avoir une connaissance assez riche du monde dont il traite. Or il n'en est rien ! Ce générateur romanesque utilise une particularité intéressante de la sémantique : à partir du moment où une séquence formelle semble bien formée dans une langue donnée, qu'elle n'est constituée que de termes de cette langue et que le lecteur accepte de

la considérer comme telle, alors il accepte d'importer en elle le sens qui ne s'y trouve pas. Par exemple, que le programme de génération ignore totalement ce que signifie « un grand ciel rouge » ne gêne en rien le lecteur qui, à partir de ses connaissances du monde, attribue un sens à cette expression. Lorsque cette règle n'est pas respectée, le fonctionnement sémantique n'est pas immédiat et demande au lecteur un plus grand effort de coopération comme dans l'extrait de poème suivant produit par le générateur de *Trois mythologies et un poète aveugle...*

la mer triture son plectre again arbres d'hiver colours in the sky glacée
par les pluies et les vents d'automne
sous le souffle vide de la mort restes de vent dans la plaine d'Ivry
lumière interne ai piedi d'una estinta cheval de pluie thick
air and wet ciel couleur de loque nuage et pluie les fleurs éclatent
comme des étoiles
compter les frissons du jour a little movement in the leaves ombre
verte pâlisante glacée par les pluies borough polluted wich provides
colours
when the poppies are out of flowers les fleurs éclatent
derrière les genévriers obscurs obscurs haies...

Cependant, le lecteur ne recule généralement pas devant cet effort pourvu qu'il accepte l'intertexte « poésie contemporaine ». En effet, à cause de l'emboîtement des contextes définissant des situations multiples de communication, il y a toujours possibilité de sens. Le problème est de savoir quel sens il y a, donc qu'est-ce qui doit être programmé pour quel usage. S'il n'y a pas possibilité d'un programme universel de traitement sémantique, existe la possibilité de nombres d'algorithmes spécifiques efficaces dédiés à des traitements particuliers. Dans certains cas limites, la production de sens peut être même entièrement abandonnée au lecteur, le programme se contentant de faire des propositions aléatoires syntaxisées dont il ne maîtrise aucune signification.

4.1) Scénarii

Ce n'est quand même généralement pas la cas. L'exemple ci-dessus du roman TRAJECTOIRES se contente de maîtriser trois types de relations...

- *La première* est l'appartenance à un univers. Cet univers est strictement défini par les possibles relationnels qu'il contient. Sa définition n'est donc pas théorique mais pragmatique. TRAJECTOIRES est un roman policier sur la terreur dont l'action se déroule à deux périodes : 1793 et 2009. Ses dictionnaires de description contiennent des représentations de connaissances adéquates à ce propos : ils ne contiennent pas la totalité des informations possibles sur 1793, mais un nombre suffisant de représentations de connaissances sur cette époque pour que le lecteur accepte cette affirmation. Par exemple, les personnages ne se déplacent pas en 1793 comme en 2009.

- *La deuxième* est le fractionnement d'un univers de connaissances en un emboîtement de micro-univers plus spécifiques et plus maîtrisables. Chacun de ces micro-univers essaie d'explorer un aspect particulier du réel avec une connaissance suffisante pour que le générateur puisse produire à son sujet une variété de textes de surface. Parmi les micro-univers, celui d'une exécution capitale en 1793, ou celui d'une soirée mondaine, ou celui de la pluie, ou celui d'une journée ensoleillée d'août. Ces micro-univers ne peuvent produire que sur un thème et un seul : le micro-univers de la pluie ne peut ainsi que dire sans arrêt « il pleut », même s'il peut le dire avec une infinie variété. Ces micro-univers sont gigognes, chacun d'entre peut contenir de nouveaux micro-univers spécifiques : le bruit de la pluie ou les cris de la populace... Ces micro-univers une fois constitués, rien n'interdit de les réemployer dans d'autres univers différents. Ainsi un univers est constitué par un ensemble de micro-univers et c'est cet ensemble qui distingue un univers d'un autre.

- *La troisième* est la scénarisation des séquences de micro-univers. Cette scénarisation dit quelle place peut, lors de la génération, occuper tel micro-univers par rapport à tel autre. Le texte ci-dessus décrivant une exécution capitale est ainsi guidé par le scénario suivant : *datation* → *arrivée des condamnés* → *montée sur l'échafaud d'un condamné* → *prise en charge du condamné par les bourreaux* → *exécution proprement dite*

Ce *scénario* constitue l'ossature de la scène où l'ordre des événements n'est pas interchangeable. Pour citer Barthes, il s'agit des *fonctions* qui structurent le récit. Cependant deux autres possibilités viennent se greffer sur cette ossature...

1. La définition d'un des micro-univers spécifiques comme une séquence de micro-univers encore plus spécifiques : l'exécution est ainsi décrite comme : *placement du condamné sur la machine* → *chute du couperet* → *chute de la tête coupée* → *saignement du corps* → *enlèvement du corps de l'échafaud*. L'intérêt de cet emboîtement de micro-univers est que le générateur peut détailler les événements. L'exécution peut être évoquée rapidement ou, au contraire, minutieusement détaillée.

2. La possibilité d'introduction, sur cette ossature, de micro-univers spécifiques non contraints par l'enchaînement des séquences. Les cris de la foule ou la météorologie, par exemple, sont de ceux-là. Le générateur peut, à tous moments, parler du temps qu'il fait, quel que soit le point de la séquence en cours auquel il est parvenu. Certains de ces micro-univers sont obligatoires, d'autres facultatifs, certains peuvent être définis comme contraints et d'autres comme aléatoires. Le programme peut ainsi générer la séquence : *météorologie* → *datation* → *météorologie* → *arrivée des condamnés* → *montée sur l'échafaud d'un condamné* → ... ou ... *datation* → *arrivée des condamnés* → *montée sur l'échafaud d'un condamné* → *prise en charge du condamné par les bourreaux* → *météorologie* → *exécution proprement dite* → ...

L'ensemble constitue un moteur de génération qui, tout en n'ayant que des connaissances limitées sur le monde, est suffisamment génératif pour, quel que soit le nombre de séquences produites, ne jamais devoir se répéter à l'identique.

4.2) Les mots et le monde

Le dernier problème qui reste à évoquer est la relation des graphes de représentation des connaissances à l'usage linguistique proprement dit. Si un graphe de relation peut « dire » qu'un homme peut sourire et que, par ailleurs, un sourire peut recevoir divers qualificatifs, ce graphe de relation ne dit rien sur la façon dont cette relation peut être

exprimée et il n’y a aucune règle formelle qui détermine cette relation à la langue.

L’homme affichait un sourire aimable
 L’homme arborait un sourire aimable
 L’homme avait un sourire affable
 L’homme montrait un [grand] sourire aimable
 L’homme souriait [avec gentillesse]
 Le sourire de l’homme était [des plus] aimables
 Un [grand] sourire [aimable] illuminait le visage de l’homme
 Un sourire [aimable] éclairait le visage de l’homme
 etc...

Si le but fixé au générateur est de produire de la variété et non un message standard, la seule solution est de recenser les modes d’expression possibles et de les codifier de façon à densifier les dictionnaires qu’ils constituent. Cela revient à dire que ces modes d’expression sont constitués de noyaux autour desquels, à des places définies, fixes ou non, peuvent venir se greffer des satellites facultatifs, noyaux et satellites étant chacun constitués de classes de termes non syntaxisés. Ce n’est en effet qu’une fois la structure fixée que la syntaxisation de surface peut se produire. Dans le cas ci-dessus, il y a recours à quatre modes d’expression différents pour exprimer un sens identiques. Et si un déplacement de la classe des qualificatifs soit sur le sourire, soit sur l’action de sourire n’est considéré que comme une variante d’un même cas, il n’y a plus alors que deux modes d’expression distincts.

La constitution des graphes de relation et des classes qui y sont attachées dépend, à la fois, des conventions acceptées par la langue et de l’inventivité de leurs auteurs. Le rôle de l’écrivain, bien souvent, est de définir de nouveaux graphes, de modifier les classes conventionnelles et de faire en sorte que ces modifications soient acceptées par le récepteur du texte. Dans ce cadre, programmer le sens consiste à constituer des dictionnaires de scénarii, de modes d’expression et de classes de termes de façon que le texte soit constitué et acceptable à l’issue de leurs parcours.

4.3) Fiction/non-fiction

Une dernière remarque : si l’homme ne disait que les déjà-là des sens possibles du monde, il lui serait loisible de n’utiliser que des modes d’expression préfabriqués, donc de disposer d’un ensemble restreint de dictionnaires à programmation relativement facile. La contrainte viendrait du monde non de la langue. Ainsi réaliser un générateur de lettre commerciale est assez facile sauf que ce générateur ne dispose d’aucune marge de liberté : si la lettre est une lettre de commande d’un livre donné, les possibilités de modes d’expression sont pauvres — à moins de vouloir en faire de la littérature — et les variations nulles car il n’est pas pensable d’inventer un titre au livre désiré. La pragmatique du monde impose ses contraintes et l’expression sémantique ne consiste qu’en leur mise en évidence. Ainsi un générateur de petites annonces n’aurait aucun sens car il ne pourrait que mettre dans une forme *élémentaire* les informations qui lui auraient été préalablement communiquées. Le sens se passe de la langue et la sémantique n’est que celle du métatexte : « *Travaux d’intérieur. Peint, carr, papier, élec, maçonnerie, plomberie. Tél. : 0145325420 ou 0676645620 Mr Bunac* »

est plus proche de la base de donnée — qui tire l’efficacité indéniable de son sémantisme de l’appartenance de ses termes à une classification reconnue — que d’un texte. Analyser le sens de cette annonces ne présente pas de difficulté majeure mais en produire ne présente aucun intérêt. À l’inverse, la fiction n’a pas à tenir compte du monde. Ce qu’elle demande, c’est la simulation d’un fonctionnement linguistique assez crédible pour que le lecteur accepte de le considérer comme vrai. Elle est davantage forme que monde. Rien de ce que dit la littérature n’est de l’ordre du réel pragmatique et si Madame Bovary est Flaubert dans ses romans, elle ne l’a justement pas été dans la réalité. Analyser le sens des fictions, dans ce cas, est presque de l’ordre de l’impossible car ce qu’il faudrait analyser ne serait guère que de l’ordre du texte à ses extérieurs. Par contre, dans la grande liberté où elle se situe par rapport au pragmatique, produire une sémantique acceptable de la fiction est beaucoup plus facile et intéressant. Car elle ne repose que sur les mécanismes d’ancrage à la production sémantique qui agit chaque locuteur d’une langue donnée.