Peut-on rire en IA ? Révision de croyances et modélisation de plaisanteries

Florence Dupin de Saint-Cyr, Henri Prade

▶ **To cite this version:**

Florence Dupin de Saint-Cyr, Henri Prade. Peut-on rire en IA ? Révision de croyances et modélisation de plaisanteries. Conférence Nationale en Intelligence Artificielle (CNIA 2019), AFIA : Association française pour l'intelligence artificielle; IRIT : Institut de Recherche en Informatique de Toulouse, Jul 2019, Toulouse, France. pp.1-9. hal-02328822

HAL Id: hal-02328822
https://hal.science/hal-02328822v1

Submitted on 23 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Peut-on rire en IA ?

Révision de croyances et modélisation de plaisanteries

Florence Dupin de Saint-Cyr
florence.bannay@irit.fr

Henri Prade
henri.prade@irit.fr

IRIT-CNRS, Université Paul Sabatier,
118, route de Narbonne, 31062 Toulouse Cedex 9

La plus perdue de toutes les journées est celle où l'on n'a pas ri.
Nicolas Chamfort*

Résumé

Il est des sujets parmi les activités cognitives humaines que l'intelligence artificielle étudie peu : comprendre le rire en est un exemple. Cet article, exploratoire et préliminaire, cherche à identifier les mécanismes à l'œuvre dans les mots d'esprit et les blagues à caractère narratif. Il semble que dans nombre d'entre eux, un mécanisme de révision de croyances opère, conduisant à une conclusion inattendue, et assurant la chute finale de la plaisanterie.

Mots Clef

Humour, révision de croyances, surprise.

Abstract

There are some human cognitive activities that have remained little studied in artifial intelligence : understanding laughter is an example. This paper, exploratory and preliminary, tries to identify the mechanisms at work in quips and narrative jokes. It seems that in many cases a belief revision mechanism is operating, leading to an unexpected conclusion, through the punchline of the jest.

Keywords

Humor, belief revision, surprise.

1 Introduction

A côté des multiples tâches spécifiques auxquelles s'est attaqué, avec succès, l'intelligence artificielle (IA) depuis ses débuts, comme la formalisation de différents types de raisonnement, la résolution de problèmes de satisfaction de contraintes, ou l'interprétation d'images, il en est d'autres qui ont été beaucoup moins étudiées, sans doute parce qu'elles correspondaient à des enjeux jugés moins importants, moins sérieux, ou qu'elles paraissaient plus délicates. Ainsi deux exemples de ces domaines particulièrement complexes et difficiles à appréhender sont l'humour et l'art. Ces deux questions présentent d'ailleurs quelques similarités. Dans les deux cas un "émetteur", le blagueur ou l'artiste présente un contenu expressif à un "récepteur", le public, supposé capable d'apprécier ce qu'on lui destine. Humour et art ont aussi en commun de souvent jouer avec, voire de bousculer, les interdits et les préjugés, au travers de leur propos. Dans les deux cas, il semble qu'il y ait un effet de surprise [25, 33] pour le public lors de la découverte de ce qui est présenté. Très schématiquement, tandis que l'art s'adresse d'abord à notre sensibilité et suscite nos émotions, il semble que la blague sollicite avant tout une appréhension logique du monde (on parle de "comprendre une plaisanterie"). De plus, la subtilité de la mécanique humoristique stimule l'intelligence et fait réfléchir. L'humour parlerait à notre intelligence réflexive, spéculative, tandis que l'art aiguillonnerait d'abord la part réactive de notre intelligence, et convoquerait davantage différentes formes de ressenti [8]. Sans doute les choses ne sont pas aussi simples, et de fait nombre d'œuvres d'art ont une dimension conceptuelle [22]. Peut-être, plus subtilement, on pourrait se risquer à dire que dans l'énonciation d'une blague on installe une situation présentant une incohérence que la chute va résoudre de manière surprenante, tandis que l'art cultiverait plus l'ambiguïté, le décalage, l'excès [36, 26]. Cependant l'humour comme l'art posent des problèmes de perception, d'analyse et de synthèse pour le public, dans les deux cas particulièrement difficiles à modéliser.

La littérature en IA sur l'humour est assez réduite. Une grande part des articles concernent la reconnaissance de situations d'ironie, ou de rire dans des textes, dans des signaux audio ou des images, voire l'évaluation d'un niveau de drôlerie, e.g., [29]. Quelques autres travaux s'intéressent à la génération de jeux de mots, ou à la dérivation d'acronymes détournés, e.g., [31]. Marvin Minsky [20] discute

*Sébastien-Roch Nicolas de Chamfort (1741-1794) Maximes, Pensées, Caractères et Anecdotes. Garnier-Flammarion, Collection GF, n°188, 1968. Il s'agit de la maxime n°80.

la théorie freudienne des plaisanteries [14], et met en évidence que au-delà des questions d'interdits, l'humour est aussi affaire de connaissance sur la connaissance, et de reconnaissance et de suppression d'incohérences.

L'ambition, modeste, de cet article est d'abord d'essayer de cerner les contours d'un domaine peu exploré, et plus particulièrement, dans une blague ou un bon mot, de comprendre et de modéliser le mécanisme provoquant l'effet susceptible de déclencher le rire. Nous nous en tiendrons à cette étape d'analyse sans considérer le problème de la synthèse de plaisanteries.

Nous passons d'abord en revue les notions liées au rire dans la section 2 pour ensuite s'intéresser à la modélisation en IA de la notion de surprise en section 3, la section 4 présente l'essentiel de notre contribution qui concerne la formalisation du mécanisme à l'œuvre dans la plupart des plaisanteries. L'article se termine par un retour sur quelques articles d'IA touchant à l'humour et conclut.

2 Les notions liées au rire

Tout le monde connait la phrase de Rabelais *Rire est le propre de l'homme* (« Mieulx est de ris que de larmes escripre, pour ce que rire est le propre de l'homme. » [24]). Rabelais s'inscrit dans la tradition d'Aristote qui souligne que le rire est une qualité spécifiquement humaine.

Selon Attardo et Raskin [5] il y a trois principales théories sur l'humour : la théorie du soulagement (l'humour aiderait à relâcher la tension nerveuse), la théorie de la supériorité (l'humour permettrait de se sentir supérieur), et la théorie de l'incongruité (l'humour consiste à désobéir aux schémas mentaux et attentes). Cette dernière théorie est la théorie dominante qui est aussi la perspective dans laquelle nous nous plaçons dans cet article (même si on peut arguer que la chute de la blague, amenant un certain retour à la cohérence, apaise une dissonance cognitive, ce qui semble aller dans le sens de la première théorie).

Beaucoup d'auteurs ont essayé de cerner les notions liées au rire telles que l'humour, le comique, l'ironie, le calembour, la plaisanterie, le cocasse, ou la dérision. Très tôt on a compilé des recueils de bons mots et de blagues, e.g., [4, 21]. [1] Ceux-ci opèrent généralement une classification thématique qui met en évidence les différentes variétés de sujet ou de registre (chez le docteur, l'argent, la religion, l'étranger, le sexe, ...) ou de forme (histoires drôles, calembours, devinettes, contrepèteries, ...) propices à l'élaboration de plaisanteries.

Certaines choses semblent être drôles par elle-même (le nez rouge du clown, une personne qui tombe, un langage amphigourique, ...), c'est la matière du comique, que met en évidence Bergson dans son célèbre livre sur le rire [7]. Son ouvrage comme son sous-titre l'indique est consacré à la signification du comique. Il identifie trois critères nécessaires au rire :

- l'objet du rire devrait se rapporter directement ou indirectement à des êtres humains ;
- le rieur devrait avoir mis de côté ses émotions ;
- le rire serait un acte social.

Selon Bergson « Est comique tout arrangement d'actes et d'événements qui nous donne, insérées l'une dans l'autre, l'illusion de la vie et la sensation nette d'un agencement mécanique ». Ainsi le rire proviendrait de l'incongruité de plaquer du mécanique sur du vivant. Bergson distingue cinq effets comiques :

- le diable à ressort : « Passons alors au théâtre. C'est par celui de Guignol que nous devons commencer. Quand le commissaire s'aventure sur la scène, il reçoit aussitôt, comme de juste, un coup de bâton qui l'assomme. Il se redresse, un second coup l'aplatit. Nouvelle récidive, nouveau châtiment. Sur le rythme uniforme du ressort qui se tend et se détend, le commissaire s'abat et se relève, tandis que le rire de l'auditoire va toujours grandissant. » ;
- le pantin à ficelles : lorsque l'on voit qu'une personne qui se croit libre est en réalité le jouet d'une autre personne ;
- l'effet boule de neige : engrenage involontaire d'actions de plus en plus énormes ;
- la répétition ;
- l'inversion : l'arroseur arrosé par exemple.

Tout ce qui précède ressort du comique qui est à distinguer de l'ironie. L'ironie est plus subtile. « L'ironie remet tout en question ; par ses interrogations indiscrètes elle ruine toute définition, dérange à tout moment la pontifiante pédanterie prête à s'installer dans une déduction satisfaite. Grâce à l'ironie, la pensée respire plus légèrement quand elle s'est reconnue, dansante et grinçante, dans le miroir de la réflexion » [16]. Elle avance souvent masquée, procède par sous-entendu, et il faut être capable de la déceler, de la reconnaître. C'est d'ailleurs un problème en Traitement Automatique des Langues dans le cadre de la reconnaissance d'opinion. Pour surmonter cette difficulté, on a pu proposer pour l'écrit la création et l'usage d'un point d'ironie : ؟ [2]. L'ironie peut se faire agressive en restant sur un mode plaisant [1], et devenir complexe dans son expression littéraire [35].

L'ironie se distingue du comique et de l'humour, dont il est maintenant question. Dans son article [25] sur les mots d'esprit, Raccah établit un parallèle entre métaphore et bon mot qui tous les deux reposent sur un effet 'manipulatoire' au plan de la communication. Il différencie le comique

1. Et les blagues d'il y a près de 2000 ans [4] font encore souvent (sou)rire : *C'est un intellectuel plein d'humour qui est à court d'argent et qui vend ses livres. Il écrit alors à son père : « J'ai une bonne nouvelle : je commence à vivre de mes livres »*, ou encore : *C'est un intellectuel qui s'est acheté un pantalon. Comme il est serré et qu'il a du mal à l'enfiler, il se fait épiler les jambes.* Le lecteur pourra à titre d'exercice formuler ces blagues dans le langage que nous proposons en Section 4. Notons que la première blague en utilisant les termes « intellectuel plein d'humour » gâche un peu son potentiel comique, et illustre déjà notre Propriété 2 puisque la surprise de sa chute est d'autant moins forte qu'on s'attend dès le début à ce que cette personne pleine d'humour en fasse preuve.

et l'humour par le fait que le premier fait rire du comportement volontaire ou non du locuteur alors que le second nécessite trois conditions :

— un énoncé ;

— une intention de faire rire ;

— que le propos cause le rire.

Raccah propose comme explication d'un bon mot qui nous fait rire « la sensation d'être tombé dans un piège inattendu et inévitable : c'est la *chute* qui nous fait tomber dans ce piège ».
Dans notre étude, nous nous restreignons à l'humour et donc nous ne traiterons ni de l'ironie ni des situations comiques comme celles étudiées par Bergson. Cependant il serait possible de retrouver des éléments de la définition de Raccah dans les situations comiques : par exemple dans la situation du diable à ressort, le spectateur peut penser que ce n'est pas possible de continuer à donner des coups de bâtons au gendarme ; la surprise, et donc le rire, vient du fait que ça ne s'arrête pas. Pour le pantin à ficelle, le spectateur peut supposer que le pantin va comprendre qu'il est manipulé, la surprise viendrait de ce qu'il ne s'en rend pas compte ...

3 Les surprises en IA

La surprise naît de la survenue de quelque chose jugée pratiquement impossible. La première version de la théorie des possibilités [10] qu'avait proposée l'économiste anglais G. L. S. Shackle [28] reposait sur la notion de degré de surprise associé à un évènement, qui était en fait un degré d'impossibilité de cet évènement, calculé à partir d'une distribution de possibilités reflétant notre connaissance incertaine de l'état du monde considéré. Un fait est alors d'autant plus surprenant qu'il est moins cohérent avec ce qu'on s'imaginait possible. Formellement,
$\texttt{degré de surprise}(A) = 1 - \texttt{possibilité}(A)$.

En IA, la question des surprises a été étudiée par Lorini et Castelfranchi [19]. Ces auteurs ont formulé deux notions de surprise différentes :

— *mismatch-based surprise* qui traduit l'incompatibilité entre ce que l'on perçoit et ce que l'on s'attendait à percevoir (*scrutinized expectation*)

— *astonishment* ou *surprise in recognition* qui traduit la difficulté d'accepter ce que l'on perçoit car cette chose perçue est très peu plausible dans l'absolu.

Dans le deuxième cas, les auteurs évoquent le fait que la personne dans cette situation peut marquer un temps d'arrêt, d'incompréhension car elle a de la difficulté à intégrer la nouvelle perception, « elle ne peut pas y croire ». Dans les deux cas le degré de surprise ou d'étonnement est associé à une probabilité. Dans le premier cas, la surprise est d'autant plus grande que la probabilité associée à l'attente initiale est élevée ; dans le second cas, l'étonnement est d'autant plus grand que la chose perçue est improbable.
Lorini et Castelfranchi proposent aussi une formalisation en logique modale avec probabilités (*logic of probabilistic quantified beliefs*) dont la sémantique est celle donnée par Fagin et Halpern [13] à laquelle un opérateur d'action standard est ajouté ainsi que des constructeurs pour parler de ce qui en train d'être examiné (en mémoire de travail : *representation under scrutiny*) et des données perçues. Ils ont modélisé les deux types de surprise dans cette logique. Ils ont ensuite proposé de modéliser l'intégration cognitive qui suit une surprise par un opérateur de changement de croyance qu'ils nomment *update process*. Ainsi chez ces auteurs la notion de surprise est d'abord définie par une forme d'inconsistance par rapport à des attentes explicites ou par rapport à des connaissances a priori. Puis une fois la surprise arrivée, l'opérateur doit mettre à jour ses connaissances.
Dans notre étude, nous proposons une approche différente puisque nous utilisons le changement de croyance pour définir la surprise (et non pour modéliser son intégration chez l'auditeur) : la nouvelle information (la chute de l'histoire) génère une révision des croyances dont le résultat est surprenant dans le sens où il est inconsistant avec ce que l'on croyait au départ. Pour plus de détails sur la révision de croyances et la mise à jour, le lecteur pourra consulter l'appendice A.

Le terme surprise est également apparu dans le cadre de l'extrapolation de scénario [11], l'objectif est de minimiser les surprises : dans le sens où une trajectoire est d'autant plus surprenante qu'elle ne respecte pas les lois statiques et dynamiques du monde étant données des observations ou des informations sur l'occurrence d'évènements à différents instants. Cette minimisation des surprises est une extension de celle introduite dans les opérateurs d'extrapolation de croyances [12] (dans un contexte plus restreint où l'on suppose que les fluents sont par défaut inertes et les surprises sont des changements imprévus de valeurs).
Notons que dans cet article nous nous restreindrons à considérer des situations statiques dans lesquels les nouvelles connaissances apprises sont surprenantes, c'est pourquoi dans cet article la notion de surprise ne correspondra pas à une évolution surprenante du monde, mais plutôt à une évolution surprenante des connaissances sur le monde. Ainsi, notre travail se situe donc dans le domaine de la révision de croyances plutôt que dans le domaine de la mise à jour [34, 18].

4 Formalisation des plaisanteries

Dans l'humour, selon Raccah [25], le piège est créé grâce au récit utilisé dans un but manipulatoire. Avec l'humour, il y a une manipulation de ce que l'on veut faire croire à l'auditeur afin qu'il tombe dans le piège. Dans cet article sur l'humour, nous proposons simplement de traduire l'analyse de Raccah en termes de révision de croyances. Ainsi, la manipulation de l'auditeur sera vue comme un processus en deux phases :

— La première phase est une révision des croyances de

l'auditeur avec des informations incomplètes destinées à suggérer une conclusion (qui s'avèrera fausse).

— La deuxième phase correspond à l'arrivée de la chute de l'histoire, qui d'une part surprend car elle est incompatible ou incongrue par rapport à la conclusion provisoire précédente, ce qui se traduira dans notre modèle par une inconsistance. D'autre part, la chute semble inéluctable ce que nous exprimons par le fait qu'elle explique logiquement les informations initiales.

Le deuxième enseignement que l'on tire de l'article de Raccah porte sur l'intensité du caractère drôle : selon Raccah, « le trait est d'autant plus drôle que le piège était inattendu et inévitable ». Cela nous incite à proposer un degré de drôlerie qui sera fonction de la différence de niveau de normalité entre les mondes obtenus dans les deux phases d'une part et de l'inéluctabilité de l'explication que fournit la chute à la situation initiale d'autre part.

4.1 Langage

On considère un langage propositionnel $\mathcal{L}$, où les propositions sont notées par des symboles grecs en minuscule ; les symboles $\bot$, $\top$, $\vee$, $\wedge$, $\neg$, $\rightarrow$, $\equiv$, $\vdash$, $\models$ dénotent respectivement la contradiction, la tautologie, les connecteurs logiques « ou », « et », « non », l'implication matérielle, l'équivalence logique, l'inférence classique, la satisfaction. On utilise $\rightsquigarrow$ pour représenter une règle par défaut : $a \rightsquigarrow b$ qui est interprétée comme la contrainte [2] $N(b|a) > 0$ signifiant qu'il est plus plausible que b soit vrai en présence de a plutôt que b soit faux [6]. Cette écriture suppose l'existence d'une relation de plausibilité sur les mondes, que nous représenterons par une distribution de possibilité π. Dans ce contexte, une base de connaissances K est un ensemble de règles par défaut $\mathcal{L} \times \mathcal{L}$ de la forme $a \rightsquigarrow b$ où $a, b \in \mathcal{L}$.
Dans la suite, nous utilisons un opérateur de *révision de croyances* [3, 17] dont les principes sont rappelés en appendice. La révision des croyances K par l'information φ est notée $K \circ \varphi$.

4.2 Formalisation

Dans cette section, nous proposons une formalisation *statique* des plaisanteries en logique propositionnelle, c'est-à-dire que l'on considère que la plaisanterie décrit une situation α et que la chute vient compléter cette description par une information β. Les informations α et β sont décrites par des propositions. Nous considérons que la plaisanterie est écoutée par un auditeur dont la base de connaissances est un ensemble de formules propositionnelles dénotée K.

Définition 1 (énoncé, énoncé à tiroir)
Une énoncé simple est une paire (α, β) *de formules de* $\mathcal{L}$. *Une énoncé à tiroir est un n-uplet avec* $n > 2$, $(\varphi_1, \ldots, \varphi_n)$, *de formules de* $\mathcal{L}$.

La situation cognitive induite par un énoncé est décrite par la base de connaissances K révisée itérativement par les éléments de l'énoncé. La chute d'une plaisanterie est considérée comme *surprenante* si le résultat de la révision des connaissances K par la description initiale de la situation est contradictoire avec ce que l'on obtient après révision par la chute.

Définition 2 (énoncé surprenant)
Un énoncé (α, β) *est* surprenant *relativement à* K *ssi*

$$(K \circ \alpha) \textit{ consistant}^{3} \textit{ et } (K \circ \alpha) \cup (K \circ (\alpha \wedge \beta)) \vdash \bot$$

Un énoncé à tiroir $(\varphi_1, \ldots, \varphi_n)$ *est surprenant à l'étape* i *relativement à* K *ssi*

$$(K \circ \psi) \cup (K \circ (\psi \wedge \varphi_i) \vdash \bot$$

avec $\psi = \varphi_1 \wedge \varphi_2 \ldots \wedge \varphi_{i-1}$ *et* $K \circ \psi$ *consistant.*

Pour comprendre la plaisanterie il faut que sa logique paraisse implacable une fois la chute révélée, c'est donc que la chute est à la fois tout à fait recevable et explique la situation. En d'autres termes, si l'on avait su β dès le début, il aurait expliqué α. Ce que l'on peut traduire ainsi :

Définition 3 (énoncé à chute révélatrice)
Étant donné un énoncé (α, β), *sa chute* β *est* révélatrice *relativement à* K *ssi*

$$(K \circ \beta) \textit{ consistant et } K \circ \beta \models \alpha$$

On peut penser que le rire vient soulager la tension de dissonance cognitive provoquée par l'inconsistance entre ce que l'on s'attendait à entendre et la chute de l'histoire. Ce soulagement ne peut avoir lieu qu'une fois l'histoire comprise, la chute jouant le rôle d'une révélation.

Définition 4 (énoncé potentiellement drôle) *Un énoncé est* potentiellement drôle *ssi il est surprenant et sa chute est révélatrice.*

Cette définition d'énoncé potentiellement drôle est directement inspirée de Raccah [25], à ceci près que nous avons essayé de définir le caractère révélateur de la chute plutôt que son côté inévitable (qui pourrait se comprendre comme $K \circ \alpha \vdash \beta$, niant le caractère surprenant de β).

Notons qu'on a écrit « potentiellement drôle » car on peut imaginer des exemples d'énoncés surprenants à chutes révélatrices mais qui ne sont pas nécessairement drôles pour leur auditeur. Par exemple, une découverte scientifique peut être surprenante et révélatrice, tout comme une

2. Cette contrainte utilise une mesure de nécessité conditionnelle. En théorie des possibilités [10], les mesures de nécessité sont duales des mesures de possibilité Π : $N(a) = 1 - \Pi(\neg a)$. La contrainte $N(b|a) > 0$ s'écrit de manière équivalente en $\Pi(a \wedge b) > \Pi(a \wedge \neg b)$ ce qui exprime que, dans le contexte a, b vrai est la situation la plus normale.

3. Cette condition est automatiquement assurée par les postulats habituels de la révision (voir Appendice) sitôt que α est consistant.

histoire policière bien ficelée, ou encore certaines révélations politiques, pour les publics concernés. Cette remarque montre que d'autres ingrédients que le mécanisme à base de révision de croyances ci-dessus sont également nécessaires à la survenue du rire. Ainsi, on peut citer les éléments évoqués en Section 2, tels que l'intention de faire rire, l'absence d'empathie, ou encore le jeu avec des tabous.

Après cette dernière définition, nous sommes en mesure de formaliser un premier exemple (issu d'un recueil d'histoires « drôles » [21]) exprimé en langage naturel que nous avons transposé en logique. Dans cet article, nous faisons abstraction de l'étape du passage de l'énoncé en langage naturel à une représentation logique.

Exemple 1 *Un homme vient de se faire renverser par une voiture. Le conducteur sort de la voiture et dit : Vous êtes bien chanceux on est juste devant le bureau du médecin. Oui ! sauf que le médecin c'est moi !*

Modélisation :
$\alpha = blesse \wedge proche_medecin$,
$\beta = blesse \wedge medecin_meme$,
$K = \left| \begin{array}{l} blesse \wedge proche_medecin \rightsquigarrow sera_vite_soigne \\ blesse \wedge \neg medecin_soignant \rightsquigarrow \neg sera_vite_soigne \\ medecin_meme \rightsquigarrow proche_medecin \\ blesse \wedge medecin_meme \rightsquigarrow \neg medecin_soignant \end{array} \right.$
Calculons maintenant $K \circ \alpha$, les modèles les plus plausibles satisfont $\varphi = \alpha \wedge sera_vite_soigne$.
Les modèles les plus plausibles de $K \circ (\alpha \wedge \beta)$ satisfont $\psi = proche_medecin \wedge medecin_meme \wedge \neg medecin_soignant \wedge \neg sera_vite_soigne$. La chute est donc bien surprenante puisque $\varphi \wedge \psi \vdash \bot$.
D'autre part, la chute explique α :

$$K \circ \beta \vdash blesse \wedge proche_medecin$$

$proche_medecin$ correspond à la partie de l'énoncé qui nous avait manipulé pour nous faire penser que le blessé était chanceux.

La propriété suivante montre l'importance de l'ordre narratif dans la plaisanterie.

Propriété 1 (ne pas réveler la chute avant la fin !) *Si (α, β) est un énoncé potentiellement drôle pour un auditeur sachant K, alors (β, α) n'est pas potentiellement drôle pour cet auditeur.*

Preuve : Puisque (α, β) est potentiellement drôle, c'est que la chute est révélatrice, c'est-à-dire que $K \circ \beta \nvdash \bot$ et $K \circ \beta \models \alpha$. Mais ceci empêche l'énoncé (β, α) d'être surprenant, et donc d'être potentiellement drôle. En effet, en utilisant les postulats K5 et K6, $K \circ (\beta \wedge \alpha) \equiv (K \circ \beta) \cup \{\alpha\}$ donc $K \circ \beta$ est consistant avec $K \circ (\beta \wedge \alpha)$. □

La propriété suivante indique la vertu de la concision.

Propriété 2 (les blagues les plus courtes...) *Si (α, β) est un énoncé potentiellement drôle sachant K alors l'ajout d'$\alpha' \in \mathcal{L}$, tel que $\alpha \nvdash \alpha'$, peut rendre l'énoncé $(\alpha \wedge \alpha', \beta)$ non potentiellement drôle.*

Preuve : En effet, soit $K = \{o \rightsquigarrow v, m \rightarrow o, m \rightsquigarrow \neg v, o \wedge a \rightsquigarrow m\}$ une base de connaissances exprimant dans l'ordre que : généralement les oiseaux (o) volent (v), les manchots (m) sont des oiseaux, généralement les manchots (m) ne volent pas, la plupart des oiseaux de l'Antarctique (a) sont des manchots. Imaginons un énoncé (o, m), il est à la fois surprenant puisque $K \circ o \models o \wedge \neg m \wedge \neg a \wedge v$ et $K \circ (o \wedge m)) \models o \wedge m \wedge \neg v$ et la chute est révélatrice puisque $K \circ m \models o$. Cependant en ajoutant l'information que l'oiseau est de l'Antarctique on peut créer une situation non surprenante vis à vis de $\beta = m$, ainsi $(o \wedge a, m)$ n'est plus un énoncé surprenant. D'autre part, si on ajoute une information qui n'a pas de rapport avec β alors on peut obtenir une chute qui n'est pas complètement révélatrice, considérons $\alpha' = t$, où t signifie « s'appelle Titi » alors $K \circ m \vdash o$ ne permet pas de déduire que $K \circ m \vdash t$ l'énoncé a donc une chute qui n'est pas nécessairement révélatrice. □

Pour se convaincre de la validité de la Propriété 2, on peut considérer le cas particulier $\alpha' = \beta$. Ceci correspondrait dans l'Exemple 1 à préciser que le conducteur vient de renverser un médecin traitant devant le cabinet de ce même médecin. Dans ce cas, la chute « Oui ! sauf que le médecin c'est moi ! » n'est plus drôle (car on le sait déjà). Néanmoins la nouvelle histoire peut être trouvée drôle différemment avec $\alpha'' =$ *Un homme vient de renverser son propre médecin traitant devant son cabinet* et $\beta'' =$ *« Vous êtes bien chanceux, on est juste devant le bureau de mon médecin »*. Ici, l'énoncé est bien surprenant cependant la révélation ne peut venir que si l'on perçoit l'ironie de l'utilisation du mot chanceux.

On peut remarquer que la Propriété 2 n'impose pas que la chute soit la plus courte possible, on peut donc affiner la définition de « drôle » en « efficace » en imposant la minimalité.

Définition 5 (concision de la chute) *Un énoncé (α, β) est efficace ssi il est potentiellement drôle et pour tout β' t.q. $\beta \models \beta'$, (α, β') n'est pas potentiellement drôle.*

Revenons sur l'Exemple 1, cet exemple peut aussi être vu comme une plaisanterie à tiroir $(\varphi_1, \varphi_2, \varphi_3, \varphi_4)$ avec $\varphi_1 = blesse$, $\varphi_2 = chanceux$, $\varphi_3 = proche_medecin$, $\varphi_4 = medecin_meme$. En effet, dire « Vous êtes bien chanceux » après avoir vu que la personne est blessé peut déjà provoquer le rire. Ensuite la révélation d'être « devant le bureau d'un médecin » permet de comprendre. Enfin la seconde partie de la plaisanterie correspond à ce qui a déjà été analysé. Notons qu'ici dans cette première partie d'énoncé itéré φ_2 est surprenant mais non révélateur, c'est φ_3 qui explique φ_2. Ainsi pour une plaisanterie à tiroir,

le caractère surprenant et le caractère révélateur ne sont pas nécessairement simultanés. En voici un autre exemple d'après [21].

Exemple 2 *Un gars et une bonne femme rentrent dans le cabinet du médecin. Celui-ci, se tournant vers la dame, lui dit :*
— Si vous êtes malade, déshabillez-vous, je vous prie... mais la fille fait des manières. Elle a des réticences. Elle regarde le gars par en-dessous. Alors le médecin répète :
— Mais enfin, madame, déshabillez-vous ! Je suis docteur. La pudeur n'est pas de mise ici ! Alors elle se met à gigoter et tout d'un coup, elle fond en larmes. Décontenancé, le médecin demande au gars :
— Qu'est-ce qu'elle a, votre femme ? Elle est toujours aussi nerveuse ?
— Je ne sais pas, dit l'autre. Je viens juste de faire sa connaissance dans votre salle d'attente....

Modélisation :
$\varphi_1 = ensemble, \quad \varphi_2 = reticences, \quad \varphi_3 = \neg couple$
$K = \left| \begin{array}{l} ensemble \rightsquigarrow couple \\ couple \rightsquigarrow \neg reticences \\ \neg couple \rightsquigarrow reticences \end{array} \right.$
Calculons maintenant $K \circ \varphi_1$*, les modèles les plus plausibles satisfont* $\psi = ensemble \wedge couple \wedge \neg reticences$*.* $K \circ (\varphi_1 \wedge \varphi_2)$ *satisfait* $reticences$*.* φ_2 *est donc surprenante puisque* $\psi \wedge (K \circ (\varphi_1 \wedge \varphi_2)) \vdash \bot$*.*
D'autre part la chute φ_3 *explique* φ_2 *:* $(K \circ \varphi_3) \models reticences$*.*

Enfin, pour donner corps à l'intuition de Raccah selon laquelle la plaisanterie est d'autant plus drôle que la surprise et le caractère inévitable sont grands, on peut proposer l'évaluation suivante :

Définition 6 (degré de drôlerie) *Le degré de drôlerie d'une plaisanterie* (α, β) *est :*

$$\max(Incons((K \circ \alpha) \cup (K \circ (\alpha \wedge \beta)), N(\alpha | K \circ \beta)))$$

Cette définition suppose que les règles par défauts de K sont codées en logique possibiliste [6], $Incons(A)$ représentant le niveau d'inconsistance de la base de connaissance possibiliste A. Le deuxième terme est la nécessité conditionnelle de α sachant $K \circ \beta$, reflètant l'inéluctabilité de la révélation.

De plus, n'oublions pas que d'autres éléments peuvent contribuer au degré de drôlerie. On peut penser à l'effort cognitif nécessaire pour comprendre la plaisanterie, le registre, les effets comiques du narrateur [4], l'humeur de l'auditeur, sa capacité d'inhibition de ses émotions...

Cette section a proposé la modélisation d'un aspect important des plaisanteries en termes de révision de croyances, illustré sur quelques exemples. Il est clair qu'une validation de la pertinence de ces idées demanderait leur vérification sur un corpus conséquent de blagues. Même si les auteurs ont l'impression que de fait beaucoup de blagues fonctionnent de cette manière.

Il faut enfin noter que d'autres mécanismes peuvent être à l'œuvre dans les blagues narratives. Ainsi un procédé qu'on rencontre relativement souvent, est la mise en parallèle de plusieurs situations, ce qui s'apparente à l'analagie, où le raisonnement n'est plus déductif, mais l'effet de surprise final toujours aussi grand. En voici un exemple [5] :
Dans le cadre de fouilles jusqu'à 100 m de profondeur, dans le sous-sol russe, les scientifiques ont trouvé des vestiges de cuivre qui dataient d'environ 1000 ans.
Les Russes en ont conclu que leurs ancêtres disposaient déjà il y a 1000 ans d'un réseau de fil de cuivre.
Les Américains, pour faire bonne mesure, ont également procédé à des fouilles dans leur sous-sol jusqu'à une profondeur de 200 m. Ils y ont trouvé des restes de verre. Il s'est avéré qu'ils avaient environ 2000 ans.
Les Américains en ont conclu que leurs ancêtres disposaient déjà il y a 2000 ans d'un réseau de fibre de verre numérique. Et cela, 1000 ans avant les Russes !
Une semaine plus tard, en Belgique on a publié le communiqué suivant : « Suite à des fouilles dans le sous-sol belge jusqu'à une profondeur de 500 m les scientifiques belges n'ont rien trouvé du tout. Ils en concluent que les Anciens Belges disposaient déjà il y a 5000 ans d'un réseau Wifi ! »

5 Autres travaux d'IA sur l'humour

Comme souligné dans le livre de Raskin [27], peu de chercheurs se sont consacrés exclusivement à la recherche sur l'humour. On peut cependant noter des publications dans les conférences d'intelligence artificielle comme celle de Hempelmann et collègues dans FLAIRS en 2006 [15] qui proposent d'utiliser les ontologies pour analyser et générer des plaisanteries automatiquement selon des scripts prédéfinis. Ces auteurs se basent sur la théorie en linguistique computationelle d'Attardo et Raskin [5] selon laquelle les six principales ressources participant à la drôlerie d'une plaisanterie sont (par ordre décroissant d'importance) : l'opposition de script (SO), le mécanisme logique (LM), la situation (SI), la cible (TA), la stratégie narrative (NS) et le langage (LA). Les auteurs choisissent ensuite de se consacrer aux deux premiers, et c'est aussi ce que nous avons proposé dans cet article puisque nous inspirant de l'article de Raccah cité plus haut, nous nous sommes basés sur la surprise et le caractère révélateur de la chute. La surprise est liée à la contradiction entre les deux phases de la plaisanterie (SO), le caractère révélateur est associé au mécanisme logique (LM).

4. En particulier, on peut citer les décalages de niveau de langage présents dans les parodies burlesques, comme dans le poème de Georges Fourest parodiant Le Cid qui se termine ainsi :

« Dieu ! » soupire à part soi la plaintive Chimène,
« qu'il est joli garçon l'assassin de Papa ! »

La Négresse Blonde (1909)

5. Tiré de `http://www.bernardwerber.com/interactif/histoiresdroles.php?page=2&tri=1`

On peut également citer les travaux de Stock et Strapparava [31] publiés à l'IJCAI en 2003 qui présentent HAHACRONYM un système de production d'acronymes humoristiques à partir d'acronymes existants (par exemple, F B I - Federal Bureau of Investigation devient Fantastic Bureau of Intimidation, ou encore MIT - Massachusetts Institute of Technology devient Mythical Institute of Theology). Notons qu'il s'agit de la génération d'un énoncé surprenant et révélateur au sens où il explique la série de lettres d'une façon inattendue.
Shahaf et collègues [29] ont travaillé sur l'analyse automatique du degré de drôlerie de la légende humoristique associée à une image, ils ont utilisé une base de milliers de légendes recueillies à l'occasion d'un concours organisé par le New York Times pour élire la légende la plus drôle associée à une même image. Les auteurs ont fait annoter les légendes par paires en demandant aux participants d'indiquer la plus drôle des deux légendes pour chaque paire. Un classifieur de légendes a ensuite été construit qui prédit étant donné un dessin et deux légendes, quelle légende est la plus drôle. Puis en simulant un tournoi le système permet d'élire la légende la plus drôle de toutes. Les caractéristiques utilisées pour formuler la tâche en termes d'apprentissage sont : le caractère inhabituel du langage utilisé, le caractère imprévisible de la légende, la complexité grammaticale. Là encore, on retrouve la nécessité du caractère surprenant pour la drôlerie.

6 Conclusion

Il arrive que certains articles ou écrits à allure scientifique, soient des plaisanteries [32, 9, 23]. De deux choses l'une, soit le présent article est une plaisanterie à propos d'articles scientifiques, soit c'est un article scientifique à propos des plaisanteries ¿! Cela dépend de la surprise que cette conclusion provoque chez le lecteur par rapport à ce qu'il s'attendait à trouver. ¿!

Appendice : Révision de croyances

Dans le cadre du changement de croyances, l'article d'Alchourròn, Gärdenfors et Makinson (AGM) [3] a introduit le concept de révision de croyances. La révision de croyances consiste à déterminer ce qu'il reste des anciennes croyances après l'arrivée d'une nouvelle information. Les croyances sont représentées par des expressions dans un langage formel. AGM ont défini trois types de changement de croyances, la contraction, l'expansion et la révision. L'expansion consiste simplement à ajouter l'information sans vérifier sa cohérence avec les connaissances antérieures. La contraction permet de retirer une information. La révision consiste à ajouter une information tout en préservant la cohérence. Cette dernière opération est nécessaire puisque l'incohérence entraîne un état de croyance inexploitable. Le principal apport de l'article d'AGM est la définition d'un ensemble de postulats qui doivent être satisfait par tout opérateur de révision dit *rationnel*. Comme remarqué dans [30] ces postulats sont fondés sur trois principes :
— un principe de consistance (le résultat doit être consistant),
— un principe de minimisation du changement (on doit modifier le moins possible les croyances initiales),
— un principe de priorité à la nouvelle information (la nouvelle information doit être vérifiée après la révision).
Nous rappelons ci-dessous l'ensemble des postulats énoncés par Katsuno et Mendelzon [18] qui sont équivalents à ceux d'AGM, mais permettent plus facilement de relier un opérateur de révision à une relation de distance entre interprétations. Plus formellement, un opérateur de révision KM associe à un ensemble de formules K et à une formule φ (représentant la nouvelle information), un autre ensemble de croyances noté $K \circ \varphi$. Pour être considéré *rationnel* l'opérateur $\circ$ doit satisfaire les postulats KM :
$K1$: $K \circ \varphi \models \varphi$
$K2$: si $K \cup \{\varphi\}$ satisfiable, alors $K \circ \varphi \equiv K \cup \{\varphi\}$.
$K3$: si φ est satisfiable, alors $K \circ \varphi$ l'est aussi.
$K4$: Si $K_1 \equiv K_2$ et $\varphi_1 \equiv \varphi_2$ alors $K_1 \circ \varphi_1 \equiv K_2 \circ \varphi_2$
$K5$: $(K \circ \varphi) \cup \{\psi\} \models K \circ (\varphi \wedge \psi)$
$K6$: si $(K \circ \varphi) \cup \{\psi\}$ est satisfiable alors $K \circ (\varphi \wedge \psi) \models (K \circ \varphi) \cup \{\psi\}$
$K1$ impose que la nouvelle information doit être vérifiée après la révision. $K2$ impose que lorsque la nouvelle information n'est pas contradictoire avec les croyances initiales alors la révision est une simple expansion. $K3$ exprime que si la nouvelle information est consistante alors l'ensemble de croyances révisées l'est aussi. $K4$ exprime qu'un opérateur de révision est indépendant de la syntaxe. Ces 4 premiers postulats sont les postulats basiques de révision, les deux derniers expriment la minimisation du changement. $K5$ implique que réviser par une conjonction $\varphi \wedge \psi$ doit donner un état de croyances plus précis que ce que l'on peut déduire logiquement de K révisé par φ auquel on a ajouté par union l'information ψ. $K6$ signifie que lorsqu'on révise K par $\varphi \wedge \psi$, toute déduction logique à partir de ψ et $K \circ \varphi$ fera partie de la croyance révisée à condition que ψ ne soit pas contradictoire avec $K \circ \varphi$.
Katsuno et Mendelzon ont énoncé le théorème de représentation suivant, permettant d'exprimer une révision en termes de modèles proches :

Théorème 1 ([17]) $\circ$ *satisfait les postulats* $(K1)$ *-* $(K6)$ *ssi il existe une fonction associant fidèlement à chaque état épistémique* φ *un pré-ordre total* $\leq_\varphi$ *tel que :*

$$Mod(\varphi \circ \psi) = \min(Mod(\psi), \leq_\varphi)$$

L'association est fidèle lorsqu'elle associe à toute formule φ un pré-ordre $\preceq_\varphi$ qui privilégie strictement les interprétations satisfaisant φ aux autres.
Dans le contexte d'une révision de croyances, l'information φ est une nouvelle connaissance sur le monde qui est considéré comme statique. Lorsque φ représente une évolution du monde et non pas une évolution des connaissances sur le monde, alors l'opération est appelée *mise à*

jour [34, 18]. La mise à jour possède ses propres postulats et son propre théorème de représentation, mais cette question va au-delà de la modélisation utilisée dans cet article.

Références

[1] René Marill Albérès. *Le Comique et l'Ironie*. Classiques Hachette, Thèmes et Parcours Littéraires, 1973.

[2] Alcanter de Brahm. *L'Ostensoir des Ironies*. Rumeur des âges (1996), précédé de Le point sur l'ironie par Pierre Schoentjes, 1899.

[3] Carlos E Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change : Partial meet contraction and revision functions. *The J. of Symbolic Logic*, 50(2) :510–530, 1985.

[4] Anonyme. *Va te marrer chez les Grecs (Philogelos)*. Traduction, notes et postface par Arnaud Zucker. Mille et une Nuits, 2008.

[5] Salvatore Attardo and Victor Raskin. Script theory revis(it)ed : Joke similarity and joke representation model. *Humor-International Journal of Humor Research*, 4(3-4) :293–348, 1991.

[6] Salem Benferhat, Didier Dubois, and Henri Prade. Practical handling of exception-tainted rules and independence information in possibilistic logic. *Appl. Intell.*, 9(2) :101–127, 1998.

[7] Henri Bergson. *Le Rire : Essai sur la Signification du Comique*. publie.net, 1900 (réédition 2011).

[8] Jean-François Bonnefon and Henri Prade. Qu'est-ce qui (nous) fait signe ? In Mario Borillo, editor, *Dans l'Atelier de l'Art : Expériences Cognitives*, pages 186–205. Champ Vallon, 2010.

[9] Miguel de Unamuno. *La Cocotologie. Notes pour un traité*. Editions Self, Paris, 1946.

[10] Didier Dubois and Henri Prade. *Possibility Theory. An Approach to Computerized Processing of Uncertainty*. Plenum Press, New York and London, 1988. With the collaboration of H. Farreny, R. Martin-Clouaire and C. Testemale.

[11] Florence Dupin de Saint-Cyr. Scenario update applied to causal reasoning. In *Proc. Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'08)*, pages 188–197. AAAI Press, septembre 2008.

[12] Florence Dupin de Saint-Cyr and Jérôme Lang. Belief extrapolation (or how to reason about observations and unpredicted change). *Artificial Intelligence*, 175 :760–790, janvier 2011.

[13] Ronald Fagin and Joseph Y Halpern. Reasoning about knowledge and probability. *J. of the ACM (JACM)*, 41(2) :340–367, 1994.

[14] Sigmund Freud. Humour. In *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XXI (1927-1931) : The Future of an Illusion, Civilization and its Discontents, and Other Works*, pages 161–166. 1961.

[15] Christian Hempelmann, Victor Raskin, and Katrina E Triezenberg. Computer, tell me a joke... but please make it funny : Computational humor with ontological semantics. In G. Sutcliffe and R.Goebel, editors, *Proc. 19th Int. Florida Artificial Intelligence Research Society Conf., FLAIRS'06*, pages 746–751. AAAI Press, 2006.

[16] Vladimir Jankélévitch. *L'Ironie*. Flammarion, 1964.

[17] Hirofumi Katsuno and Alberto O Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52(3) :263–294, 1991.

[18] Hirofumi Katsuno and Alberto O Mendelzon. On the difference between updating a knowledge base and revising it. *Belief revision*, 29 :183, 2003.

[19] Emiliano Lorini and Cristiano Castelfranchi. The cognitive structure of surprise : looking for basic principles. *Topoi*, 26(1) :133–149, 2007.

[20] Marvin Minsky. Jokes and the logic of the cognitive unconscious. In Lucia Vaina and Jaakko Hintikka, editors, *Cognitive Constraints on Communication - Representations and Processes*, pages 175–200. D. Reidel Publ. Comp., 1984.

[21] Hervé Nègre. *Dictionnaire des Histoires Drôles*. Fayard, 1970.

[22] Henri Prade. Monde(s) et représentation(s). quelques propos sur les rapports entre art et science. In Mario Borillo, editor, *Approches Cognitives de la Création Artistique*, pages 49–56. Mardaga, 2005.

[23] Georges Pérec. *Cantatrix sopranica L. et autres écrits scientifiques*. Éditions du Seuil, coll. « Librairie du XXe siècle », 1980 (réédité en 1995).

[24] François Rabelais. *La vie très horrificque du grand Gargantua, père de Pantagruel, jadis composée par M. Alcofribas abstracteur de quintessence*. La Sirène (réédition 1919), 1534.

[25] Pierre-Yves Raccah. Humour et métaphore : quelques éléments d'une analogie pour la construction d'un sens inattendu : Illustration sur un corpus de citations de George Bernard Shaw. *Revue de Sémantique et Pragmatique, Presses de l'Univ. d'Orléans*, 2015.

[26] Vilayanur S. Ramachandran. The artful brain. In *The Internet and the University : Forum 2004*, pages 169–198. 2004. https ://www.educause.edu/ir/library/pdf/ffpiu049.pdf.

[27] Victor Raskin. *The Primer of Humor Research*, volume 8. Walter de Gruyter, 2008.

[28] George Lennox Sharman Shackle. *Decision, Order and Time in Human Affairs*. (2nd edition), Cambridge University Press, UK, 1961.

[29] Dafna Shahaf, Eric Horvitz, and Robert Mankoff. Inside jokes : Identifying humorous cartoon captions. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1065–1074. ACM, 2015.

[30] Léa Sombé. A glance at revision and updating in knowledge bases. *International J. of Intelligent Systems*, 9(1) :1–27, 1994.

[31] Oliviero Stock and Carlo Strapparava. Getting serious about the development of computational humor. In *Proc. IJCAI'03*, volume 3, pages 59–64, 2003.

[32] Thémiseul de Saint-Hyacinthe. *Le Chef d' Œuvre d'un Inconnu*. Réédition Aubanel (1965), 1714.

[33] M.-A. Williams. Aesthetics and the explication of surprise. *Languages of Design. Formalisms for word, image and sound*, 3 :145–157, 1996.

[34] Marianne Winslett. Reasoning about action using a possible models approach. In *Proc. AAAI'88*, page 89–93, 1988.

[35] Monique Yaari. *Ironie Paradoxale et Ironie Poétique*. Summa Publications, Birmingham, Alabama, 1988.

[36] Semir Zeki. *Inner Vision. An Exploration of Art and the Brain*. Oxford University Press, 1999.