HAL open science

La compréhension des histoires drôles : une affaire de révision de croyances

Florence Dupin de Saint-Cyr, Henri Prade

To cite this version:

Florence Dupin de Saint-Cyr, Henri Prade. La compréhension des histoires drôles : une affaire de révision de croyances. Revue Ouverte d'Intelligence Artificielle, 2022, Post-actes de la Conférence Nationale en Intelligence Artificielle (CNIA 2018-2020), 3 (3-4), pp.253-285. 10.5802/roia.31 . hal-03694646

HAL Id: hal-03694646

https://hal.science/hal-03694646v1

Submitted on 13 Jun 2022

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.

Distributed under a Creative Commons Attribution 4.0 International License

Revue Ouverte d'Intelligence Artificielle

Florence Dupin de Saint-Cyr, Henri Prade

La compréhension des histoires drôles : une affaire de révision de croyances

Volume 3, nº 3-4 (2022), p. 253-285.

<http://roia.centre-mersenne.org/item?id=ROIA_2022__3_3-4_253_0>

© Association pour la diffusion de la recherche francophone en intelligence artificielle et les auteurs, 2022, certains droits réservés.

Cet article est diffusé sous la licence
Creative Commons Attribution 4.0 International License.
http://creativecommons.org/licenses/by/4.0/

La Revue Ouverte d'Intelligence Artificielle est membre du Centre Mersenne pour l'édition scientifique ouverte
www.centre-mersenne.org

Revue Ouverte d'Intelligence Artificielle
Volume 3, nº 3-4, 2022, 253-285

La compréhension des histoires drôles : une affaire de révision de croyances

Florence Dupin de Saint-Cyr[a], Henri Prade[a]

[a] IRIT-CNRS Université Paul Sabatier 118, route de Narbonne 31062 Toulouse Cedex 9, France
E-mail : florence.bannay@irit.fr, henri.prade@irit.fr.

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 cherche à identifier les mécanismes à l'œuvre dans les mots d'esprit et les blagues à caractère narratif et logique. 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. Les blagues à fonctionnement analogique sont aussi évoquées. L'article comporte aussi une substantielle étude bibliographique couvrant les travaux anciens ou moderne sur le rire (illustrée par quelques bons mots).

Mots-clés. — Humour, révision de croyances, surprise, incohérence, incongruité.

La plus perdue de toutes les journées
est celle où l'on n'a pas ri.
Nicolas Chamfort[(1)]

1. Introduction[(2)]

À côté des multiples tâches spécifiques auxquelles s'est attaquée, 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.

[(1)] 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.

[(2)] Cet article est une version révisée (sic) et augmentée de deux articles récents de conférences [30, 31]. Il contient en particulier une étude bibliographique importante sur les nombreuses modélisations informelles sur le rire (qui était très réduite dans ces deux articles).

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 [66, 122] (Latour 1956, Williams 1996)[3] 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 [17] (Bonnefon et Prade 2010). Sans doute les choses ne sont pas aussi simples, et de fait nombre d'œuvres d'art ont une dimension conceptuelle [89] (Prade 2005). 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 [92, 125] (Zeki 1999, Ramachandran 2004). 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.

De plus, il est clair que le rire faisant partie de la vie ordinaire des gens cela en fait un sujet important et légitime de recherche dans les différentes disciplines concernées. Tout le monde connaît 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. » [90] (*Gargantua* 1534)). Rabelais s'inscrit dans la tradition d'Aristote qui souligne que le rire est une qualité spécifiquement humaine. Pour Baudelaire, en 1868, dans *De l'essence du rire* [10], « Le rire est satanique, il est donc profondément humain. Il est dans l'homme la conséquence de l'idée de sa propre supériorité ».

Mais le rire a de nombreuses autres facettes et a intéressé des chercheurs dans les différentes disciplines scientifiques[4] touchant à l'humain [110] (Smadja 1993), depuis la physiologie [95] (Raulin 1900), jusqu'à la philosophie, la psychologie, la linguistique, l'anthropologie, l'ethnologie, la sociologie. La psychanalyse a mis en évidence l'importance du rapport entre l'inconscient et le rire, notamment en lien avec les failles du système défensif [19] (Bourgain 2010). Il est clair aussi que l'histoire, la littérature, les beaux-arts et la caricature sont également concernés. On pourra consulter parmi les ouvrages récents [70] (Letourneux et Vaillant 2021) pour un vaste panorama transdisciplinaire, allant du XIXe à nos jours, traitant des différents aspects

[3] Dans cet article, pour permettre au lecteur de situer rapidement les références, la plupart d'entre elles sont suivies du nom des auteurs et de la date de parution.

[4] On notera l'existence de plusieurs journaux uniquement consacrés à ce sujet : Humor : Int. J. of Humor Research (depuis 1988), European J. of Humor Research, (depuis 2013), Israeli J. of Humor Research : An international journal (depuis 2012), Japanese J. of Laughter and Humor Research, (depuis 1994).

du rire en occident, ou encore [76] (Melchior-Bonnet 2021) pour l'histoire de la place du rire des femmes dans la société.

Cet article laisse de côté la plupart de ces aspects, mis à part les éléments de modélisation de ce qui provoque le rire apportés par la philosophie et la psychologie, qui sont rappelés dans la Section 2. Dans la suite, on se focalise exclusivement sur les mécanismes du rire à l'œuvre quand on raconte des histoires drôles. L'ambition, modeste, de cet article est d'abord d'essayer de cerner les contours d'un domaine peu exploré en IA, et plus particulièrement, dans une blague ou un bon mot, de comprendre et de modéliser ce qui est susceptible de déclencher le rire. Nous nous en tiendrons à cette étape d'analyse sans considérer le problème de la synthèse de plaisanteries. On exclut également de l'étude, le rire provoqué par les dessins humoristiques, ou les situations grotesques qui mettent en jeu non seulement des mécanismes de compréhension, mais aussi de perception et d'analyse de scènes, statiques ou dynamiques.

Ainsi, nous dressons d'abord un panorama de l'abondante littérature sur le rire, en philosophie puis en psychologie, et enfin en linguistique et IA. La section 3 rappelle d'abord les notions de révision de croyances (sous-section 3.1) qui seront utiles dans la formalisation des plaisanteries proposée dans cet article, puis passe en revue la modélisation de l'idée de surprise en sous-section 3.2. La section 4 présente l'essentiel de notre contribution qui concerne la formalisation du mécanisme à l'œuvre dans la plupart des plaisanteries à caractère logique. Enfin, les blagues à fonctionnement analogique sont brièvement discutées.

2. Panorama de travaux sur l'humour en rapport avec l'IA

Cette section passe d'abord en revue des études concernant des théories sur le rire développées, pour la plupart, par des philosophes et des psychologues, pour ensuite rappeler quelques travaux existants en IA.

2.1. Généralités sur le rire

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. Le premier livre en langue française sur ce sujet, en 1579, est sans doute celui de Laurent Joubert [55], médecin du roi. Loin d'être découragé par la difficulté de la tâche, il se lance avec enthousiasme dans une étude de près de 350 pages de ce mystère merveilleux qui lui semble une des plus « amirables accions » de l'homme. À partir de la fin du 18e siècle et plus particulièrement depuis les 150 dernières années, des monographies sont consacrées au rire et à l'humour : depuis James Beattie en 1776 [11] et ses successeurs [15, 52, 66, 72, 82, 112, 114] (Hunt 1848, Spencer 1860, Bergson 1900, Sully 1902, Ludovici 1932, Monro 1951, Latour 1956). Un ouvrage ambitieux récent [53] (Hurley 2011) propose « d'expliquer pourquoi l'humour existe, comment ça fonctionne dans le cerveau et pourquoi la comédie est un art ».

Parmi les travaux présentant un panorama des recherches sur le rire, celui de Patricia Keith-Speigel en 1972 [62] est de grande ampleur bibliographique et souvent

cité. Elle propose un panorama des différentes théories sur l'humour qui ont émergé au 19e siècle et pendant la première moitié du 20e, en précisant toutefois que certaines théories « compliquent plus qu'elles ne démêlent ce nœud gordien ». Nous rappelons brièvement ces théories en les séparant un peu arbitrairement entre philosophie et psychologie.[(5)]

2.2. Travaux en philosophie

Trois idées ont été souvent mises en avant comme à l'œuvre dans le déclenchement du rire : le sentiment de supériorité, l'incongruité et la surprise.

- *Supériorité* : Selon F. R. Fleet en 1890 [35], se référant au *Leviathan* [48] (Hobbes 1651), le rire ('laughter') « is a sudden glory arising from some sudden conception of some eminency in ourselves by comparison with the inferiority of others ». Ainsi ce qui est risible serait la mise en avant d'une imperfection (absente chez l'auditeur) par celui qui veut faire rire, excitant un sentiment de supériorité. Les auteurs de cette théorie voient le rire comme une victoire sur d'autres personnes (ou circonstances). On se sent supérieur en riant car on est moins bête, moins moche, moins malchanceux, moins faible que la cible de la moquerie ou du ridicule.
- *Incongruité* : Beattie [11] dès 1776, relève que l'humour peut être basé sur les idées/situations mal appairées ou divergentes par rapport aux coutumes habituelles, il emploie même l'expression « situations incohérentes ou inappropriées ». Kant en 1790 [57], comme Schopenhauer en 1819 [104] parle de l'incongruité soudaine de la perception d'un objet réel comparée à ce qui avait été imaginé/attendu. Spencer en 1860 [112] reprend cette idée tout en soulignant que l'incongruité n'est pas nécessairement source de rire puisqu'elle peut aussi provoquer l'émerveillement. L'étude publiée en 1900 par Bergson [15], sur laquelle nous reviendrons plus en détails, avec son expression « du mécanique plaqué sur du vivant » a aussi sa place dans cette catégorie. Des auteurs variés accompagnent l'incongruité d'un autre ingrédient, comme Guthrie en 1903 [44] qui souligne que l'incongruité ne sera risible que dans la mesure où on est assuré par ailleurs que « tout va bien » tandis que Delage en 1919 [24] insiste plutôt sur la nécessité de la présence d'un désagrément pour autrui : « Pour être comique, une chose doit réunir deux conditions essentielles : i) Il doit exister entre l'effet produit et sa cause une désharmonie quantitative ou qualitative, d'où résulte une impression de surprise, un effet d'imprévu ; ii) l'effet doit être désagréable pour la personne qui le subit. » On peut trouver d'autres références et plus de détails sur la théorie de l'incongruité dans l'ouvrage très complet de Chapman et Foot paru en 1976 [22].
- *Surprise* : La surprise, qui comme on l'a vu n'est pas sans rapport avec le point précédent, est considérée comme une condition nécessaire au rire par

[(5)] Une étude plus récente par une chercheuse en linguistique computationnelle [78] (Mihalcea 2007) présente une vue d'ensemble multidisciplinaire des travaux sur l'humour verbal, incluant aussi les premiers travaux en génération et en reconnaissance d'expressions à caractère humoristique.

beaucoup d'auteurs, depuis Descartes en 1649 [25] qui considère que le rire résulte d'un mélange de joie peu intense et de choc, en passant par l'étude psychologique faite par [51] (Hollingworth 1911) qui montre que lorsque la nouveauté ou la surprise est éliminée alors la réaction à l'humour est diminuée. Notons que Marius Latour [66] a dédié le quatrième essai de son livre de 1956 à « la surprise source du rire ».

Ces trois notions se retrouvent dans le célèbre livre de Bergson sur le rire [15]. Pour cet auteur, certaines choses semblent être drôles par elles-mêmes (le nez rouge du clown, une personne qui tombe, un langage amphigourique, ...), c'est la matière du comique. 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 » [54] (Jankelevitch 1964). Elle avance souvent masquée, procède par sous-entendu, et il faut être capable de la déceler, de la reconnaître.

Dans notre approche présentée en Section 4, nous nous restreignons à l'humour dans les plaisanteries narratives et donc nous ne traiterons ni de l'ironie ni des situations comiques comme celles étudiées par Bergson.

2.3. Travaux en psychologie et sciences humaines

En ce que concerne le rire, les études en psychologie tiennent une place de plus en plus importante en particulier depuis cinquante ans. Dans cette section nous continuons de parcourir la vue d'ensemble donnée par Patricia Keith-Speigel [62], sur le versant psychologie et sciences humaines avant de nous intéresser à des travaux plus récents. Voici les cinq théories qui complètent selon cette chercheuse les trois catégories dégagées par les philosophes (supériorité, incongruité, surprise).

- *Théories biologiques et évolutionnistes* : Darwin en 1890 [23] a étudié le rire comme manifestation particulière d'émotions et plutôt du point de vue phénoménologique : « Joy, when intense, leads to various purposeless movements – to dancing about, clapping the hands, stamping, etc., and to loud laughter. Laughter seems primarily to be the expression of mere joy or happiness. We clearly see this in children at play, who are almost incessantly laughing. » Il est reconnu que le rire et l'humour sont des constituants du système nerveux et ont des fonctions adaptatives (bon pour le corps, répondant à un instinct, moyen de communication antérieur à l'usage de la langue, moyen social pour exprimer l'unité d'un groupe et maintenir ses normes, relique de l'attitude de combat (dents, secousses)). Des travaux comportementaux comme [33] (Epstein et Joker 2007) montrent notamment qu'il existe un seuil de durée optimale entre la délivrance du contexte d'une histoire et l'arrivée de la chute pour que l'histoire soit perçue comme drôle. Nous laissons ici de côté les aspects physiologiques, éthologiques, anthropologiques du rire (voir par exemple [110] (Smadja 1993)), ainsi que des travaux récents en neurosciences comportementales [121] (Vrticka et col. 2013).
- *Théories de l'ambivalence* : Elles reposent sur l'idée que le rire résulte du fait que le récepteur ressent simultanément des émotions incompatibles. En particulier Socrate explique le rire par le ressenti simultané du plaisir et de la douleur dans un des dialogues de Platon (*Le Philèbe* ou *Sur le plaisir*). Cette idée se trouve aussi dans le livre déjà cité de Laurent Joubert [55].
- *Théories de la libération et du soulagement* : Rejoignant les physiologistes, [41, 64, 112] (Spencer 1860, Kline 1907, Gregory 1924) considèrent que la fonction première de l'humour serait de soulager la tension en excès. Il s'agit ici de la tension qui peut exister dans l'auditoire préalablement à la décision d'un locuteur de raconter une blague. Nous évoquerons plus loin un autre type de tension, celle induite à l'audition de la chute d'une blague.
- *Théories configurationnelles* : L'idée à la base de ces approches est que des éléments perçus initialement comme sans rapports s'organisent soudainement ensemble. À l'inverse de la théorie d'incongruité qui met en avant la discordance, les théories configurationnelles mettent en avant le rapprochement. Ces théories sont les prémisses de la psychologie de la forme (théorie de la Gestalt) mise en évidence dans le cadre de l'humour par

Maier en 1932 [73]. Dans son analyse Maier met en avant des éléments comme la surprise, la double interprétabilité, l'absence de compassion, le rapprochement et l'incongruité[(6)]. Dans le même ordre d'idée et préfigurant une problématique d'intelligence artificielle, Schiller [103] en 1938 dans sa *théorie de la dualité dynamique* voit les plaisanteries comme une variété de problème à résoudre, ce point de vue est partagé par Bateson [9] en 1953 qui décrit le mécanisme de compréhension d'une blague comme celui utilisé pour la perception du fond et de la forme (*figure-ground perception*).

- *Théories psychanalytiques* : Freud [37] en 1905 dit que quand l'énergie d'un canal psychique n'est pas ou ne peut pas être utilisée (à cause de la censure du super ego) alors cette énergie peut être déchargée avec plaisir par le rire, ce qui est lié à la théorie du soulagement. Freud voit dans le rire une expression du besoin humain de minimiser le déplaisir et maximiser le plaisir. Ainsi pour lui, le comique court-circuite (donc économise) le raisonnement, le mot d'esprit permet de contourner des inhibitions, l'humour permet d'éviter la compassion. Dans un article ultérieur de 1928, Freud [36] élabore le « triomphe du narcissisme », à l'œuvre dans l'humour, rejoignant le principe de supériorité, où « non seulement le moi triomphe, mais aussi le principe de plaisir qui est assez fort pour s'affirmer face au caractère défavorable des circonstances réelles. »

Depuis les années 1970, de multiples ouvrages académiques sont dédiés à cette thématique : on peut se reporter par exemple à la série de livres en psychologie qui ont fait le point à différents moments sur la théorisation du rire et de l'humour [22, 40, 74, 100] (Goldstein et McGhee 1972, Chapman et Foot 1976, Roeckelein 2002, Martin et Ford 2018).

Attardons-nous un peu plus particulièrement sur les travaux de 1983 d'une psychologue française, Françoise Bariaud, sur la génèse de l'humour chez le jeune enfant [8]. Son étude s'appuie sur la réception par de jeunes enfants de dessins humoristiques qui ont l'avantage d'être compréhensible sans savoir lire. Parmi les explications de l'effet comique avancées par Françoise Bariaud nous relevons que

- la notion d'incongruité intervient en accord avec les théories psychologiques vues précédemment. Pour ce faire, la notion de référentiel cognitif (que l'enfant est en train de construire) qui sert à la fois à détecter l'incongruité et à « la résoudre » [sic]. La distance par rapport au référentiel cognitif est importante car au delà d'une certaine distance, le stimulus peut être considéré comme absurde alors qu'en deçà la compréhension (donc la « résolution de

[(6)] « The thought-configurations which makes for a humorous experience must (1) be unprepared for ; (2) appear suddenly and bring with it a change in the meaning of its elements ; (3) be made up of elements which are experienced entirely objectively... ; (4) contain as its elements the facts appearing in the story, and these facts must be harmonized, explained and unified ; and (5) have the characteristics of the ridiculous in that its harmony and logic apply only to its own elements » [73] [pp.73-74].

l'incongruité ») est possible. Selon Françoise Bariaud, la « résolution de l'incongruité » fait appel à un sens logique. Elle note aussi que la mise en jeu de tabou n'est pas nécessaire.

- la différence entre deux modes : le mode ludique et le mode sérieux. Lors de l'écoute d'une histoire, si on est en mode sérieux alors on se prend au jeu de la situation, il y a de la compassion pour les héros comme si c'était réel alors que dans le mode ludique on sait que c'est faux et on met l'accent sur le côté irréel. L'entrée dans le mode ludique peut se faire juste avant que l'histoire commence, on attend une histoire drôle cela déclenche une ouverture à quelque chose d'inattendu et à une certaine complicité (il faut pour rire une volonté d'entrer dans le jeu ou un désir de complicité avec l'émetteur du stimulus humoristique). Notons encore que pour que le récepteur de la blague s'installe dans le mode ludique, il faut que l'incongru ne mette pas en cause sa sécurité.
- la notion de difficulté de résolution comme partie prenante du plaisir à comprendre : la nécessité d'un effort cognitif (associé à une tension) produit quand il est terminé un soulagement sous la forme d'un sourire ou d'un rire, elle se réfère à l'indicateur du plaisir de maîtrise cognitive chez le jeune enfant introduit par Kagan en 1974 [56].

En résumé, les travaux de Françoise Bariaud sur la fabrication psychique et psychoaffective du rire conduisent à dire que « sur le plan cognitif, l'identification d'une incongruité ou d'une absurdité associé à un effet de surprise va faire l'objet d'un traitement spécifique dont le résultat positif accordera un qualificatif dit risible, comique à la représentation traitée victorieusement » [110] (Smadja 1993).

2.4. Vers des approches IA de l'humour

Un des fondateurs de l'IA est aussi un des premiers à discuter des plaisanteries en termes de représentation et raisonnement sur les connaissances. Il s'agit de Marvin Minsky [81] qui discute en 1984 la théorie freudienne des plaisanteries [37] (Freud 1905) 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. Minsky est un des premiers en IA à formaliser la notion de double entente en humour, il le fait au moyen de sa théorie des frames « The elements that seems to me most common to all the different kinds of humor is that of unexpected frame-substitution, in which a scene is first described from one viewpoint and then suddenly – typically by a single word – one is made to view all the scene-elements in another, quite different way ». Citons aussi [77] (Mele 2002) qui s'intéresse à la modélisation du ridicule en se basant sur la théorie des attitudes mentales de Shoham et Cousins de 1994 [107]. Il définit le ridicule réel ou fictionnel comme provenant nécessairement d'une action inadéquate (d'aucuns diraient « incongrue »), la condition suffisante au ridicule serait basée sur les concepts d'attente et de surprise.

La littérature en IA sur l'humour est cependant 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. Quelques autres travaux s'intéressent à la génération de jeux de mots, ou à la dérivation d'acronymes détournés. Ce sont ces travaux que nous évoquons brièvement avant d'aborder des recherches autour de la formalisation linguistique de l'humour, puis des approches cognitives plus proches de l'IA.

2.4.1. *Détection de l'ironie et de l'humour, génération de formules humoristiques*

Comme nous l'avons vu, l'ironie est une forme d'humour qui repose sur une expression à double entente (pas toujours mise en évidence). L'ironie peut se faire agressive en restant sur un mode plaisant [2] (Alberes 1973), et devenir complexe dans son expression littéraire [124] (Yaari 1988). Elle n'est pas toujours facile à détecter. Cela a contribué au succès des émoticônes 🤔 et autres smileys ;-) pour lever toute ambiguïté. Mais l'idée du point d'ironie est ancienne puisqu'il y a plus d'un siècle, le poète Alcanter de Brahm [20] en avait proposé en 1899 un graphisme élégant : ⸮.

Certains travaux d'IA relevant de l'humour se sont attaqués au problème de la détection de l'ironie en traitement automatique des langues dans le cadre de la reconnaissance d'opinion [7, 59] (Kaboui et col. 2019, Barbieri et col. 2015) ou encore, en communication personne-machine, l'identification de l'ironie dans des expressions faciales [1] (Albanie et Vedaldi 2016) ou non-verbales [116] (Tahon et Devillers 2015), la génération de l'expression corporelle du rire dans les agents conversationnels animés [84, 85] (Ochs et col. 2017, Niewiadomski et col. 2010).

Dès 2006, l'apprentissage automatique supervisé est utilisé pour identifier si des textes sont humoristiques ou non [79] (Mihalcea et Strapparava 2006). D'autres travaux, utilisent des classificateurs bayésiens ou des SVMs [21] (Buscaldi et Rosso 2007). Toujours en apprentissage, [109] (Sjöberg et Araki 2007) procèdent à des classifications en s'appuyant sur la sélection de caractéristiques saillantes comme la similarité textuelle (avec la blague la plus proche), la présence de mots drôles, l'ambiguïté des mots utilisés (évaluée à partir du nombre de sens de chaque mot, donnés par un dictionnaire), le style (évalué en vérifiant par la prononciation donnée dans le dictionnaire, la présence de rimes ou d'allitérations), la présence d'expressions idiomatiques (vérifiée par l'appartenance de groupes de mots à une liste de proverbes). Dans [80] (Mihalcea et col. 2010), les auteurs testent si une chute terminant un contexte est ou non comique (compris au sens d'incongru), en se basant sur des représentations vectorielles de textes et des mesures appropriées aux blagues (polysémie, allitération, similarité avec les blagues du dataset). Récemment [46] (Hasan et col. 2021) se sont attaqués (par des techniques d'apprentissage avancé basées sur des architectures de type « Transformer ») à la détection multimodale de l'humour dans des vidéos, ce qui requiert la prise en compte des composantes verbales et non verbales ainsi que l'intégration de connaissances externes.

Shahaf et collègues [106] ont travaillé en 2015 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.

Un autre domaine un peu plus confidentiel est la génération automatique de formules humoristiques, dans ce registre, on peut citer les travaux de Stock et Strapparava [113] publiés à l'IJCAI dès 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 M.I.T. – Massachusetts Institute of Technology devient Mythical Institute of Theology). Notons qu'il s'agit de la génération d'un énoncé surprenant mais logique au sens où il explique la série de lettres d'une façon inattendue et même incongrue[(7)].

D'autres auteurs [68] (Lessard et Levison 1992) ont développé un système automatique de générations d'expressions humoristiques appelé *Tom Swifties* qui utilise des schémas simples en exploitant des relations de synonymie, méronymie, antonymie, ainsi que d'homophonie par exemple : « You must be my host, Tom guessed ». Leur système a ensuite été prolongé avec de nouveaux schémas [69] (Lessard et col. 2002) permettant de générer des bons mots basés sur des homonymies comme : « Joan kisses a hero at the disco. A knight club ».

Comme on l'a vu plus haut, certains bons mots reposent sur la mise en place d'une divergence de sens en utilisant des mots homophones. Kao *et col.* [58] sont parmi les premiers en 2016 dans le cadre d'un modèle probabiliste de la langue à définir les notions d'ambiguité et distinguabilité entre deux phrases qui diffèrent par un mot phonétiquement ambigu, en utilisant respectivement l'entropie et la divergence de Kullback–Leibler. Leurs expérimentations montrent que les phrases jugées les plus drôles par 100 participants sont corrélées à la fois à une forte ambiguïté et une forte distinguabilité. Par exemple, la phrase « A dentist has to tell a patient the whole tooth. » est identifiée comme drôle tandis que « A dentist examines one tooth at a time » ne l'est pas.

2.4.2. *Théorisations psychologiques ou linguistiques de l'humour*

L'un des premiers modèles proposé en linguistique pour l'analyse des plaisanteries est celui de Suls en 1972 [115]. C'est un modèle en deux phases : la chute de la plaisanterie qui est incongrue et inattendue, suivie d'une résolution (pensée en termes

[(7)] Dans le même ordre idée, on peut mentionner la génération d'anagrammes de mots ou de groupes de mots qui semblent révélateurs d'un sens caché. Ainsi « Le romantisme » donne « Larmoiements », ou encore « George Sand et Alfred de Musset » donne « Le fumet des grands orages d'été » [88] (Perry-Salkow et Castelain 2020) et « Albert Einstein » donne « Rien n'est établi » [63] (Klein et Perry-Salkow, 2011).

de résolution de problème) qui restaure la cohérence. Plus précisément, ce modèle suggère quatre facteurs qui peuvent contribuer à l'appréciation de l'humour par le récepteur :

(1) l'incongruité de la chute, c'est-à-dire à quel point la chute viole les attentes (ce que Françoise Bariaud considèrera plus tard comme une distance avec le référentiel cognitif du récepteur),
(2) la complexité de la phase de résolution. Il y a une mobilisation cognitive pour résoudre l'incongruité : une blague trop facile ou trop dure ne sera pas drôle. Cette notion est liée au principe de congruence cognitive de [126] (Zeigler et col. 1967), « dans les expériences effectuées par ces auteurs, l'appréciation s'est avérée en effet d'autant plus forte que le stimulus était Â« en accord avec la complexité des processus cognitifs du sujet Â», autrement dit requérait de lui un certain effort intellectuel (qui ne doit pas être non plus trop intense) » [8] (Bariaud 1983).
(3) le temps pris à résoudre l'incongruité,
(4) la saillance du contenu de la blague.

Dans le même ordre d'idée, Shultz [108] qui s'intéresse au développement cognitif de l'appréciation des devinettes chez l'enfant, souligne en 1976 à la suite d'autres auteurs que l'incongruité est en général définie comme un conflit entre ce qui est attendu et ce qui se produit vraiment dans la blague.

Notons que Suls propose de traiter de la même façon une blague et un dessin humoristique avec légende : « It is suggested here that a joke or cartoon is found to be funny as the result of a two-stage process. In the first stage, the perceiver finds his expectations about the text disconfirmed by the ending of the joke or, in the case of a cartoon, his expectations about the picture disconfirmed by the caption » [115] (Suls 1972).

Ces considérations peuvent faire penser aux maximes conversationnelles de Grice [42] (Grice 1975) qui régissent la bonne coopération dans un dialogue entre le locuteur et son interlocuteur. De fait, Attardo dans le chapitre 9 de son ouvrage de 1994 [4] explique que les plaisanteries violent les maximes de Grice (celles de quantité, de manière, parfois de qualité, mais remarquablement pas celle de relation[(8)]), voir aussi [32] (Dynel 2008). Notons que [93] (Attardo et Raskin 1994) définissent de nouvelles maximes de communication *non-bona fide* spécifiques du cadre humoristique.

Dès 1991, Attardo et Raskin [6] retiennent principalement trois théories sur l'humour parmi celles rappelées plus haut : la théorie du soulagement (l'humour aiderait

[(8)] Grice's maxims :

- quantity : make your contribution as informative as required,
- manner : avoid obscurity of expression, avoid ambiguity, be brief (avoid unnecessary prolixity), be orderly.
- quality : do not say what you believe to be false ; do not say that for which you lack adequate evidence.
- relation : be relevant.

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 du soulagement). Ces auteurs, revisitant des idées antérieures de Raskin de 1984 [94] basées sur l'usage de scripts, introduisent la théorie GTVH (General Theory of Verbal Humor) [6] qui met en avant six principales ressources participant à la drôlerie d'une plaisanterie (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). Ils ont étudié les différences et similarités entre blagues en faisant varier le script de la blague selon ces six caractéristiques et ont testé leurs effets humoristiques sur des sujets humains [101] (Ruch et col. 1993). Dans ce cadre, [47] (Hempelmann et col. 2006) utilisent des ontologies pour analyser et générer des plaisanteries automatiquement à partir de scripts.

Il apparaît sur les expérimentations de [101] (Ruch et col. 1993) que (SO) et (LM) sont les deux caractéristiques les plus importantes dans GTVH. C'est d'ailleurs sur le principe (SO) que se base l'approche de Tinholt et Nijholt de 2007 [118, 119] qui analyse l'humour des phrases contenant des anaphores ambiguës comme par exemple : « Les policiers ont arrêté les manifestants parce qu'ils étaient violents. » (Qui ? les policiers ?) Cette analyse se base sur la représentation des phrases anaphoriques sous forme de graphe. Selon ces auteurs, les deux scripts associables à une anaphore ambiguë ont des graphes qui se recouvrent forcément et pour que cette anaphore soit drôle, il doit y avoir une opposition entre les deux candidats à la résolution de l'anaphore « ils » (ici, les policiers ou les manifestants), cette opposition est quantifiée en utilisant les antonymies et synonymies entre les propriétés qui leurs sont associées (humains, vivants, groupés, ordonnés comparé à humains, vivants, groupés, tapageurs). Leur système est implémenté dans un agent conversationnel animé car, selon Tinholt et Nijholt, « l'humour joue un rôle important dans les conversations humaines. Les gens utilisent l'humour pour rendre une conversation agréable, pour se détendre au cours d'une discussion et pour dissimuler les erreurs et les malentendus ».

Les principes (SO) et (LM) sont aussi mis en avant (en 2015) chez Raccah [91] qui propose une analyse des mots d'esprit basée sur la surprise et le caractère révélateur de la chute. Ainsi, la surprise est liée à la contradiction entre les deux phases de la plaisanterie (SO), le caractère révélateur est associé au mécanisme logique (LM). Plus précisément, 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 et l'humour[(9)] par le fait que le premier fait rire du comportement volontaire ou non du locuteur alors que le second nécessite trois conditions :

[(9)] Bien que dédiée à l'étude des plaisanteries, il serait possible de retrouver des éléments de la définition de Raccah dans les situations comiques décrites par Bergson : 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

- 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 ».

2.4.3. *Travaux de formalisation des histoires drôles*

Graene Ritchie [99] (Ritchie 2018) est un des rares chercheurs actifs à la frontière de la psychologie et de l'IA qui, comme dit dans [74] (Martin et Ford 2018), « a suggéré que les recherches en IA sur l'humour peuvent non seulement aider les théories sur l'humour mais aussi conduire à d'importantes découvertes sur l'intelligence humaine, le langage, la résolution de problèmes, et plus généralement le traitement de l'information ».

Il s'est d'abord intéressé (avec Kim Binsted en 1997) à la génération automatique de devinettes jouant sur les mots, à partir de schémas sémantiques et syntaxiques, et a développé le système JAPE (Joke Analysis and Production Engine) [16] capable de produire par exemple : « What do you get when you cross breakfast food with a murderer ? A cereal killer. » ou encore « What's the difference between leaves and a car ? One you brush and rake, the other you rush and brake ». Le système est basé sur un ensemble de règles symboliques mélant des combinaisons sémantiques et des expressions syntaxiques. Ces règles sont utilisées dans un programme qui accède à un vaste dictionnaire avec des informations phonétiques sur l'usage lexical des mots. Wiscraic (Witty Idiomatic Sentence Creation Revealing Ambiguity In Context) est un système développé dans [75] (McKay 2002), semblable à beaucoup d'égards à JAPE, mais ayant comme objectif la génération de traits d'esprits en utilisant des expressions idiomatiques.

Une motivation avancée par Binsted et Ritchie [16] pour ce travail de génération de bons mots est qu'expérimenter leurs mécanismes de construction est une première étape dans l'analyse de l'humour. D'ailleurs, McKay [75] souligne deux aspects de l'humour présent dans les bons mots : « A pun consists of using a word in a manner that suggests two or more of its meanings, or the meaning of at least one of its homonyms ». Ainsi la transformation homonymique peut-être vue comme la chute du bon mot. De même, dans notre traitement des histoires drôles, on verra la présence de deux interprétations concurrentes du contexte, tandis que la chute fera émerger l'interprétation à laquelle on n'avait pas pensé spontanément.

Pour passer à l'analyse de l'humour, [96, 97] (Ritchie 1999, 2002) aborde la résolution des incongruités en travaillant sur des devinettes incongrues du style « Why do birds fly south in winter ? It's too far to walk ». Ritchie identifie la présence de trois entités dans de telles devinettes :

peut supposer que le pantin va comprendre qu'il est manipulé, la surprise viendrait de ce qu'il ne s'en rend pas compte.

- M1 : la première interprétation (la plus évidente) qu'on fait du contexte
- M2 : l'interprétation du contexte à laquelle on ne pense pas en premier (sens caché)
- M3 : l'interprétation de la chute (qui servira à privilégier M2 plutôt que M1)
- M4 : l'interprétation formée par l'intégration de la chute avec M2.

Ainsi dans l'exemple ci-dessus on pense d'abord (M1) que la question porte sur la destination des oiseaux, alors qu'une interprétation non évidente de la question (M2) pourraient être pourquoi ils volent, en lisant la chute (M3) on comprend (M4) qu'on répond à la question « pourquoi ils volent » et de manière incongrue. Notons que dans d'autres devinettes l'incongruité peut apparaître dès le contexte (« Postmaster : Here's your five-cent stamp. Shopper : Do I have to stick it on myself ? Postmaster : Nope. On the envelope. »). Cette façon de raisonner est appelée désambiguation de surprise. Il énonce 5 propriétés d'intérêt dans ce cadre :

- Évidence (Obviousness) : M1 est plus susceptible que M2 d'être invoquée par le lecteur ;
- Conflit : M3 ne fait pas sens avec M1 ;
- Compatibilité : M3 fait sens avec M2 ;
- Contraste : il y a des différences marquées entre M1 et M2 (possiblement M1 et M4) ;
- Inadéquation (Inapropriateness) : M4 est intrinsèquement étrange, absurde, excentrique, grotesque, ou encore tabou.

Un facteur qui en découle est la notion de surprise (ou d'attente déçue) qui est produite par le Contraste entre ce qui est prédit par l'Évidence et déçu par l'arrivée de la chute. Notons que Ritchie souligne que pour produire une résolution d'incongruité en deux étapes à la Suls [115] (Suls 1972), la présence d'une ambiguïté est nécessaire dans le contexte initial de la blague.

Ritchie [97] introduit un langage logique pour formaliser l'interprétation des blagues, il part du principe que « la chute force à réinterpréter le contexte précédemment mis en place ». Ce langage comporte un opérateur dit d'accommodation (servant à intégrer des connaissances non conflictuelles) et un opérateur dit de « mise à jour de croyances » (belief update) pour prendre en compte la chute. L'idée de l'opérateur de mise à jour est inspirée par la révision au sens de [38] (Gärdenfors 1988).

La formalisation proposée par Ritchie est novatrice par l'introduction explicite d'un opérateur de changement de croyances. Cependant, il utilise sans nécessité un opérateur d'accommodation (que la révision de croyances prend déjà en compte, puisque l'expansion [38] est une forme de révision). Il ne fait référence qu'à ses cinq propriétés basées sur M1 à M4, mais ne précise pas l'opérateur de révision utilisable qui satisferait ces propriétés, ni ne fait référence aux postulats AGM [3] (Alchourrón et col. 1985, voir Section 3.1). Au plan de la formalisation, le travail donne une impression d'inachevé, resté au niveau de définitions de base. Il ne semble pas que Ritchie ait jamais continué ce travail de formalisation. En 2014, Ritchie [98] poursuit pourtant son étude sur un plan un peu plus informel en affirmant que depuis

longtemps il y a un consensus sur l'existence d'une logique dubitative à l'œuvre dans les blagues (*dubious logic*). Il fait alors la distinction entre la *logique interne* de la blague (connue seulement du narrateur) et la *pseudo-logique* qui est utilisée par l'audience pour résoudre l'incongruité. Ces deux logiques sont illustrées par la blague traditionnelle [117] (Tibbals 2000) se moquant des irlandais : « Savez-vous ce qui est arrivés aux scouts marins irlandais ? Leur tente a coulé ». Selon Ritchie la logique interne de cette blague est : *Les scouts montent des tentes. Par conséquent, les scouts marins monteront des tentes. Les scouts marins pratiquent des activités sur la mer. Par conséquent, les scouts marins monteront des tentes sur la mer. Une tente montée en mer coulera.*, tandis que l'inférence faite par l'audience serait de la forme : *La tente a coulé, elle devait donc se trouver sur une étendue d'eau. C'est parce qu'elle appartenait à un groupe lié à la mer. Ils ont monté une tente parce qu'ils sont scouts, et les scouts montent des tentes.* Il conclut « It is appropriate, therefore, to investigate in more detail what such a 'pseudo-logic' might consist of » sans réussir à en cerner les contours (l'article ne comporte aucune référence en logique, a fortiori en logique non-classique). Sa monographie plus récente [99] (Ritchie 2018), bien que très complète à beaucoup d'égards, n'apporte pas de nouveaux éléments sur ce point.

Alors que nous parvenons au bout de notre vaste panorama bibliographique, il nous apparaît qu'il y a souvent une confusion latente entre incohérence et incongruité (en ce que ce dernier terme renvoie à l'idée de violation de normes collectives). Nous pensons que l'incongruité est un ingrédient supplémentaire qui par sa nature comique facilite le rire, et s'ajoute donc à la surprise liée au contraste (voire à l'incohérence) de la chute avec l'attente créée par le contexte.

3. Changement de croyances et surprises

Nous abandonnons provisoirement la thématique du rire pour rappeler dans cette section des éléments de formalisation en IA de deux notions clés pour notre analyse de la compréhension des plaisanteries : la révision de croyances et la notion de surprise.

3.1. Révision de croyances

Dans le cadre du changement de croyances, l'article d'Alchourròn, Gärdenfors et Makinson (AGM) [3] a introduit en 1985 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 satisfaits par tout opérateur de révision dit *rationnel*. Comme remarqué dans [111] (Léa Sombé 1994) ces postulats sont fondés sur trois principes :

- un principe de cohérence (le résultat doit être cohérent),
- 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 en 1991 par Katsuno et Mendelzon dans [61] 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 une formule[(10)] κ et à une formule φ (représentant la nouvelle information), un autre ensemble de croyances noté $\kappa \circ \varphi$. Pour être considéré *rationnel* l'opérateur $\circ$ doit satisfaire les postulats KM :

(KM1) : $\kappa \circ \varphi \models \varphi$

(KM2) : si $\kappa \cup \{\varphi\}$ satisfiable, alors $\kappa \circ \varphi \equiv \kappa \cup \{\varphi\}$

(KM3) : si φ est satisfiable, alors $\kappa \circ \varphi$ l'est aussi

(KM4) : Si $\kappa_1 \equiv \kappa_2$ et $\varphi_1 \equiv \varphi_2$ alors $\kappa_1 \circ \varphi_1 \equiv \kappa_2 \circ \varphi_2$

(KM5) : $(\kappa \circ \varphi) \cup \{\psi\} \models \kappa \circ (\varphi \wedge \psi)$

(KM6) : si $(\kappa \circ \varphi) \cup \{\psi\}$ est satisfiable alors $\kappa \circ (\varphi \wedge \psi) \models (\kappa \circ \varphi) \cup \{\psi\}$

(KM1) impose que la nouvelle information doit être vérifiée après la révision. **(KM2)** impose que lorsque la nouvelle information n'est pas contradictoire avec les croyances initiales alors la révision est une simple expansion. **(KM3)** exprime que si la nouvelle information est cohérente alors l'ensemble de croyances révisées l'est aussi. **(KM4)** 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. **(KM5)** 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 κ révisé par φ auquel on a ajouté par union l'information ψ. **(KM6)** signifie que lorsqu'on révise κ par $\varphi \wedge \psi$, toute déduction logique à partir de ψ et $\kappa \circ \varphi$ fera partie de la croyance révisée à condition que ψ ne soit pas contradictoire avec $\kappa \circ \varphi$.

Katsuno et Mendelzon ont énoncé le théorème de représentation suivant, permettant d'exprimer une révision en termes de modèles proches :

Théorème 3.1 ([61]). — *$\circ$ satisfait les postulats ((KM1))–((KM6)) si et seulement si il existe une fonction assignant fidèlement à chaque état épistémique κ un pré-ordre total $\leqslant_\kappa$ tel que :*

$$\mathrm{Mod}(\kappa \circ \varphi) = \min\,(\mathrm{Mod}(\varphi), \leqslant_\kappa)$$

(10) Dans l'approche initiale de AGM, un ensemble déductivement clos K de formules était considéré pour représenter les connaissances initiales, KM ont montré que l'on pouvait énoncer les postulats sur une formule κ dont l'ensemble des conséquences forment K.

L'assignation est fidèle lorsqu'elle associe à toute formule κ un pré-ordre $\preceq_\kappa$ qui privilégie strictement les interprétations satisfaisant κ aux autres. Plus formellement, l'assignation est fidèle ssi pour toute formule κ, on a :

1) si $\omega \models \kappa$ et $\omega' \models \kappa$ alors $\omega =_\kappa \omega'$.
2) si $\omega \models \kappa$ et $\omega' \not\models \kappa$ alors $\omega \prec_\kappa \omega'$.
3) si $\kappa \equiv \kappa'$ alors $\preceq_\kappa = \preceq_{\kappa'}$.

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* [60, 123] (Winslett 1988, Katsuno et Mendelzon 1991). 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.

3.2. Les surprises en IA

La surprise naît de la survenue de quelque chose jugée pratiquement impossible. La première version de la théorie des possibilités [26] (Dubois et Prade 1988) qu'avait proposée l'économiste anglais G. L. S. Shackle en 1961 [105] 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 imaginait possible. Formellement,

$$\texttt{degré de surprise}(A) = 1 - \texttt{possibilité}(A).$$

En IA, la question des surprises a été étudiée par [71] (Lorini et Castelfranchi 2007). 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 en 1994 par Fagin et Halpern [34] 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'incohérence 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 incohérent avec ce que l'on croyait au départ.

Le terme surprise est également apparu dans le cadre de l'extrapolation de scénarios où l'objectif est de minimiser les surprises [28] (Dupin de Saint-Cyr 2008) : 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 [29] (Dupin de Saint-Cyr et Lang 2011) (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 lesquelles 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*.

4. Vers une formalisation des plaisanteries

Dans l'humour, selon [91] (Raccah 2015), 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 cette idée en termes de révision de croyances comme l'a aussi suggéré Ritchie [97] en 2002[(11)]. 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).

[(11)]Les versions préliminaires de cet article [30, 31] étaient inspirées des idées informelles de [91] (Raccah 2015). Il apparait rétrospectivement que le modèle proposé prolonge la tentative de modélisation de [97] (Ritchie 2002).

- 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 incohérence. 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 modèles 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.

Par ailleurs, on s'en tient dans cet article à l'étude du raisonnement mené par un agent à l'écoute d'une plaisanterie. Nous laissons de côté le raisonnement qu'a pu effectuer le locuteur pour la mettre au point en l'adaptant à ce qu'il croit savoir de l'auditoire, sans parler de l'évaluation de l'opportunité de raconter ou non la blague dans le contexte.

4.1. Syntaxe et sémantique du langage utilisé

Nous considérons un langage propositionnel $\mathcal{L}$, où les propositions sont notées par des symboles grecs en minuscules. Les symboles $\perp$, $\vee$, $\wedge$, $\neg$, $\rightarrow$, $\equiv$, $\models$ dénotent respectivement la contradiction, les connecteurs logiques « ou », « et », « non », l'implication matérielle, l'équivalence logique, la satisfaisabilité. Dans ce qui suit, nous utilisons un opérateur $\circ$ de *révision de croyance* [3, 61], voir la sous-section 3.1.

Notons que la sémantique de la révision des croyances est basée sur les systèmes de sphères [43] (Grove 1988) et les enracinements épistémiques [38] (Gärdenfors 1988), deux notions qui ont leur contrepartie exacte dans la théorie des possibilités (sous la forme de distributions de possibilité qualitative et de relations de nécessité [27] (Dubois et Prade 1991)). Une distribution de possibilité est qualitative dans le sens où elle ne joue que le rôle d'une fonction de classement ordinal sur l'ensemble des interprétations possibles (contrairement à ce que ferait une distribution de probabilité). Les distributions de possibilités sont utilisées ici pour implémenter l'opérateur de révision $\circ$ car ce qui est requis est simplement un ordre de plausibilité comme dans les systèmes de sphères et aucune valeur numérique n'est ni nécessaire ni disponible. Cependant, l'approche présentée dans cet article est valable pour tout type d'opérateur de révision de croyance satisfaisant les postulats de Katsuno–Mendelzon [61].

Puisque nous devons encoder des connaissances de sens commun, nous utilisons $\rightsquigarrow$ pour représenter une règle par défaut : $\alpha \rightsquigarrow \beta$, avec $\alpha, \beta \in \mathcal{L}$; ce qui signifie que lorsque α est vrai, il est plus plausible que β soit vrai que faux. Ce type de règle se traduit facilement dans le cadre de la théorie des possibilités [26] (Dubois et Prade 1988) par la contrainte $N(\beta|\alpha) > 0$. En raison de la dualité entre une mesure de nécessité N et sa mesure de possibilité associée Π, c'est-à-dire, $N(\alpha) = 1 - \Pi(\neg\alpha)$, la contrainte $N(\beta|\alpha) > 0$ s'écrit de manière équivalente comme $\Pi(\alpha \wedge \beta) > \Pi(\alpha \wedge \neg\beta)$ qui exprime

que, dans le contexte α, avoir β vrai est la situation normale (puisque strictement plus possible que β faux). Cette écriture suppose l'existence d'une relation de plausibilité sur les interprétations, qui est représentée par une distribution de possibilités π, c'est-à-dire une fonction de l'ensemble des interprétations de $\mathcal{L}$ vers une échelle linéairement ordonnée bornée par 0 et 1, où 0 représente l'impossible et 1 le tout à fait possible.

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 des propositions de $\mathcal{L}$.

Nous considérons que la plaisanterie est écoutée par un auditeur dont la base de connaissances est un ensemble de formules propositionnelles dénotée K. Lors d'une révision de croyances, K sera assimilée à la conjonction de ses formules (notée κ dans le rappel de la section précédente). De plus l'auditeur est aussi caractérisé par la façon $\circ$ dont il révise ses croyances.

Dans la suite, une plaisanterie (simple) est vue comme un énoncé qui est une paire (<contexte>, <chute>) ayant des caractéristiques particulières. Nous définissons un énoncé comme suit :

Définition 4.1. — *Un* énoncé simple *est une paire* (α, β) *de formules de* $\mathcal{L}$. *Un* énoncé à tiroir *est un n-uplet* $(\varphi_1, \ldots, \varphi_n)$, *avec* $n > 2$, *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 4.2. — *Un énoncé* (α, β) *est* surprenant *relativement à un auditeur associé à* $(K, \circ)$ *ssi*

$$(K \circ \alpha) \text{ cohérent}^{(12)} \text{ et } (K \circ \alpha) \cup (K \circ (\alpha \wedge \beta)) \models \bot$$

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

$$(K \circ \psi) \text{ cohérent et } (K \circ \psi) \cup (K \circ (\psi \wedge \varphi_i)) \models \bot$$

avec $\psi = \varphi_1 \wedge \varphi_2 \ldots \wedge \varphi_{i-1}$.

Notons que la définition ci-dessus exclut la possibilité que l'auditeur connaisse déjà la blague. En effet, dans ce dernier cas, α et β seraient déjà déductibles de K, interdisant

(12) Cette condition est automatiquement assurée par les postulats habituels de la révision (voir 3.1) sitôt que α est cohérent.

toute surprise. Par ailleurs, nous n'exigeons pas que $K \circ (\alpha \wedge \beta)$ soit cohérent, incluant les chutes absurdes dans la définition 4.2[13].

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 4.3. — *Étant donné un énoncé* (α, β)*, sa chute* β *est* révélatrice *relativement à* $(K, \circ)$ *si*

$$(K \circ \beta) \text{ cohérent}^{(13)} \text{ et } K \circ \beta \models \alpha$$

On peut penser que le rire vient soulager la tension de dissonance cognitive provoquée par l'incohérence 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 4.4. — *Un énoncé est* potentiellement drôle *ssi il est surprenant et sa chute est révélatrice.*

Cette définition d'énoncé potentiellement drôle est directement inspirée de [91] (Raccah 2015), 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 $K \circ \alpha \models \beta$, niant le caractère surprenant de β).

Observons également que la définition repose à la fois sur K et $\circ$, qui peuvent différer d'un auditeur à l'autre, ce qui est en accord avec le fait qu'une blague peut ne pas être trouvée drôle par tout le monde.

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 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 » [83] (Negre 1970)) 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.

---

[13] Cependant, les chutes absurdes ne seront pas révélatrices au sens de la prochaine définition 4.3 en raison des postulats (KM5) et (KM6) :
(KM5) : $(K \circ \varphi) \cup \{\psi\} \models K \circ (\varphi \wedge \psi)$; (KM6) : si $(K \circ \varphi) \cup \{\psi\}$ est cohérent alors $K \circ (\varphi \wedge \psi) \models (K \circ \varphi) \cup \{\psi\}$.

[13] Cette condition est automatiquement assurée par les postulats habituels de la révision sitôt que β est cohérent, voir Section 3.1.

Dans notre traitement des exemples, nous prenons en compte le fait que l'auditeur utilise des connaissances par défaut pour comprendre une histoire. Cette façon de coder les connaissances *non-monotones* de l'agent n'enlève rien à la généralité de notre approche. C'est simplement une façon compacte et explicite de spécifier les relations de plausibilité entre interprétations (qui selon [39] (Gärdenfors 1990) sous-tendent tout opérateur de révision $\circ$). Ainsi, la connaissance de l'auditeur est exprimée par une base de connaissances cohérente[14] $\Sigma = P \cup \Delta$ qui est l'union d'un ensemble P de formules propositionnelles et un ensemble Δ de règles « défaisables » de la forme $\alpha \rightsquigarrow \beta$ où $\alpha, \beta \in \mathcal{L}$. Ici, Σ joue à la fois le rôle de K et de $\circ$ puisque Δ permet d'induire un ordre préférentiel sur les interprétations. Plus précisément, à partir de l'ensemble Σ, on peut calculer [12] (Benferhat et col. 1997) une distribution de possibilité π_Σ sur les interprétations qui satisfont les contraintes induites par Δ et attribuer 0 à chaque interprétation ne satisfaisant pas P. Par souci de brièveté, dans les exemples, tout le mécanisme de révision est caché derrière la notation « $\Sigma \circ \alpha$ », qui est calculée par la conditionnalisation de π_Σ par α (à savoir $\pi_\Sigma(\omega|\alpha) = \pi_\Sigma(\omega)$ si $\omega \models \alpha$, et 0 sinon), ce qui donne un opérateur de révision satisfaisant les postulats de la révision [14] (Benferhat et col. 2002). Cet opérateur est similaire à celui de [67] (Lehmann 1995).

Exemple 4.5. — Un homme vient de se faire renverser par une voiture. Le conducteur sort de la voiture et dit : « Vous avez de la chance, nous sommes juste en face d'un cabinet médical ». Oui ! Sauf que le docteur, c'est moi !

Modélisation :

$$\alpha = blesse \wedge procheMedecin$$
$$\beta = blesse \wedge medecinLuiMeme$$

Nous supposons que l'auditeur possède les connaissances suivantes :

$$\Sigma = \left| \begin{array}{l} blesse \wedge procheMedecin \rightsquigarrow seraViteSoigne \\ blesse \wedge \neg medecinSoignant \rightsquigarrow \neg seraViteSoigne \\ medecinLuiMeme \rightsquigarrow procheMedecin \\ blesse \wedge medecinLuiMeme \rightsquigarrow \neg medecinSoignant \end{array} \right.$$

En raison du postulat (KM1) [61] $\Sigma \circ \alpha \models \alpha$. De plus, les modèles les plus plausibles de $\Sigma \circ \alpha$ satisfont $\varphi = \alpha \wedge seraViteSoigne$.

Les modèles les plus plausibles de $\Sigma \circ (\alpha \wedge \beta)$ satisfont $\psi = procheMedecin \wedge medecinLuiMeme \wedge \neg medecinSoignant \wedge \neg seraViteSoigne$. La chute est surprenante puisque $\varphi \wedge \psi \models \bot$.

Par ailleurs, la chute explique α : $\Sigma \circ \beta \models blesse \wedge procheMedecin$. En effet, *procheMedecin* est la partie de l'histoire qui nous a manipulés en nous faisant croire que le blessé avait de la chance.

[14] La cohérence d'une base de défauts équivaut en pratique à l'existence d'une stratification non triviale [12]. Comme c'est en général le cas en logique, une base de connaissances incohérente traduit une impossibilité de raisonnement rationnel, donc ici l'impossibilité de comprendre la blague.

Notons que l'exemple 4.5 peut également être vu comme un énoncé à tiroir $(\varphi_1, \varphi_2, \varphi_3, \varphi_4)$ avec $\varphi_1 = blesse$, $\varphi_2 = chanceux$, $\varphi_3 = procheMedecin$, $\varphi_4 = medecinLuiMeme$. En effet, dire « Vous avez beaucoup de chance » après avoir constaté que la personne est blessée peut déjà provoquer le rire. Ensuite, la révélation d'être « devant le cabinet d'un médecin » aide à comprendre. Enfin, la deuxième partie de la blague correspond à ce qui a déjà été analysé. Notons que dans la première partie de l'histoire en cascade, φ_2 est surprenant mais pas révélateur, c'est φ_3 qui explique φ_2. Donc pour un énoncé à tiroir, les effets de surprise et de révélation ne sont pas forcément simultanés.

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

Proposition 4.6 (ne pas révéler 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.*

Démonstration. — Puisque (α, β) est potentiellement drôle, c'est que la chute est révélatrice, c'est-à-dire que $K \circ \beta \not\models \bot$ 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 (KM5) et (KM6)[14], $K \circ (\beta \wedge \alpha) \equiv (K \circ \beta) \cup \{\alpha\}$ donc $K \circ \beta$ est cohérent avec $K \circ (\beta \wedge \alpha)$. □

La propriété suivante indique la vertu de la concision.

Proposition 4.7 (les blagues les plus courtes...). — *Si (α, β) est un énoncé potentiellement drôle sachant K alors l'ajout de $\alpha' \in \mathcal{L}$, tel que $\alpha \not\models \alpha'$, peut rendre l'énoncé $(\alpha \wedge \alpha', \beta)$ non potentiellement drôle.*

Démonstration. — En effet voici un contre-exemple. 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. Un autre contre-exemple consiste à ajouter 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 $K \circ m \models o$ ne permet pas de déduire que $K \circ m \models t$ l'énoncé a donc une chute qui n'est pas nécessairement révélatrice. □

Pour se convaincre de la validité de la Propriété 4.7, on peut considérer le cas particulier $\alpha' = \beta$. Ceci correspondrait dans l'exemple 4.5 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 4.7 n'impose pas que la chute soit la plus courte possible. La définition suivante affine la définition de « drôle » en « efficace » en imposant la minimalité (en termes de modèles).

Définition 4.8 (concision de la chute). — *Soit* $\alpha, \beta, \beta' \in \mathcal{L}$ *et* $\beta \not\equiv \beta'$, *si* $K \cup \{\beta\} \models \beta'$ *alors l'énoncé* (α, β) *est dit plus efficace que l'énoncé* (α, β') *pour un auditeur sachant* K.

En d'autre termes, la chute est dite plus efficace si elle est moins précise et donc moins spécifique. Voici un autre exemple issu du recueil de Nègre [83].

Exemple 4.9. — 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 :

$\varphi_1 = ensemble, \quad \varphi_2 = reticences, \quad \varphi_3 = \neg seConnaissent$

$$\Sigma = \left| \begin{array}{ll} P & \neg seConnaissent \rightarrow \neg couple \\ \hline \Delta & \begin{array}{l} ensemble \rightsquigarrow couple \\ couple \rightsquigarrow \neg reticences \\ \neg couple \rightsquigarrow reticences \end{array} \end{array} \right.$$

Calculons maintenant $\Sigma \circ \varphi_1$, les modèles les plus plausibles satisfont :

$$\psi = ensemble \wedge couple \wedge \neg reticences$$

$\Sigma \circ (\varphi_1 \wedge \varphi_2) \models reticences$. φ_2 est donc surprenante puisque $\psi \wedge (\Sigma \circ (\varphi_1 \wedge \varphi_2)) \models \bot$.

De plus, la chute φ_3 explique φ_2 : $(\Sigma \circ \varphi_3) \models reticences$.

Supposons maintenant que le type ait répondu« Désolé, nous ne sommes pas en couple car on ne se connait pas » : $\varphi'_3 = \neg couple$. Il est clair que $P \cup \{\varphi_3\} \models \varphi'_3$. Par conséquent, la nouvelle chute φ'_3 est moins efficace.

Remarquons que nous aurions pu envisager la révision itérée [65] (Konieczny et Pino Pérez 2000) et l'expansion abductive [86] (Pagnucco 1996). Cependant, afin de capturer les aspects cognitifs de la compréhension d'une blague, l'idée principale est de sélectionner le meilleur état épistémique plutôt que de maintenir à jour l'ordre des

préférences sur tous les états comme dans la révision itérée. En effet, dans la révision de base, les états épistémiques peuvent être radicalement modifiés. La surprise ne serait pas la même si l'auditeur avait à l'esprit toutes les explications potentielles. À cet égard, un argument similaire s'applique à l'expansion abductive dont le but est de fournir des raisons pour le nouvel état épistémique. En outre, nous avons privilégié la simplicité de la théorie classique de la révision.

On pourrait aussi penser à utiliser le cadre de la révision à crédibilité limitée initialement introduit pas [45] (Hansson et al. 2001) et développé par [18] (Booth 2014), dans laquelle certaines informations sont considérées crédibles et d'autres pas. Dans le premier cas, la révision peut se faire sans problème (typiquement les révisions par le contexte de l'histoire, noté α dans notre approche). Dans l'autre cas, certaines informations sont non-crédibles (cela pourrait par exemple se produire avec la chute β de l'histoire) : l'auditeur n'intègre pas l'information (la révision n'est pas effectuée) mais la plausibilité de celle-ci augmente [18]. La prise en compte d'informations non-crédibles nous placerait dans un cadre plus large où l'on pourrait considérer que la chute d'une histoire est quelquefois inacceptable pour l'auditeur (par exemple, trop tirée par les cheveux).

4.3. Gradualité dans la drôlerie

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 la définition de deux quantités pour classer les plaisanteries. Ce qui permettra de construire un ordre partiel sur les plaisanteries destinées à même auditeur. Cette définition suppose que les règles par défauts Δ de Σ sont codées en logique possibiliste [13] (Benferhat et col. 1998), avec $Incons(A)$ représentant le niveau d'incohérence de la base de connaissances possibiliste A. Le deuxième terme est la nécessité conditionnelle de α sachant $\Sigma \circ \beta$, reflètant l'inéluctabilité de la révélation.

Définition 4.10 (ingrédients pour l'humour). — *Soit une base* $\Sigma = \Delta \cup P$*, un énoncé* (α, β) *est associé aux degrés de possibilités suivants :*

- *un degré de surprise :* $Incons((\Sigma \circ \alpha) \cup (\Sigma \circ (\alpha \wedge \beta)))$
- *un degré de révélation :* $N(\alpha|\Sigma \circ \beta)$

De plus, n'oublions pas que d'autres éléments peuvent contribuer à la drôlerie, comme peut en témoigner la littérature passée en revue, notamment l'effort cognitif nécessaire pour comprendre la plaisanterie, le registre, les effets comiques du narrateur[15], l'humeur de l'auditeur, sa capacité d'inhibition de ses émotions ... Tout cela

[15] 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)

participe à l'incongruité qui plus généralement fait appel à la *violation de normes* ce qui peut provoquer un effet comique (dans la mesure ou la violation n'est pas trop choquante pour le récepteur). Dans la littérature, l'expression « résoudre une incongruité » revient souvent, elle signifie trouver une logique ou tout au moins trouver la chose initiale moins incongrue qu'elle n'y paraissait, c'est-à-dire « la rendre acceptable ».

Comme nous l'avons dit précédemment nous pensons important de distinguer incongruité et incohérence. Le degré d'incongruité ne correspond pas forcément à un degré de surprise logique, tel qu'on l'a présenté dans notre formalisation. L'incongruité est plutôt comme une épice, qu'on choisit de manière plus ou moins appropriée, qu'on ajoute en plus ou moins grande quantité à un plat selon les convives auxquels il est servi. Ainsi, pour certains auditoires, certaines blagues seront trop salées, ou trop fades.

5. Une autre sorte de blague basée sur l'analogie

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'analogie, où le raisonnement n'est plus déductif, mais l'effet de surprise final toujours aussi grand.

La phrase de [81] (Minsky 1984) déjà rappelée plus haut : « The element that seems to me most common to all the different kinds of humor is that of unexpected frame-substitution, in which a scene is first described from one viewpoint and then suddenly – typically by a single word – one is made to view all the scene-elements in another, quite different way. », invoque l'idée que ce changement de « frame » peut-être dû à une analogie, une analogie *fautive* du genre « A's are like B's, except for D ». On peut en trouver des exemples dans [5, 6] (Attardo 2002, Attardo et Raskin 1991), où un glissement de « script » analogique est à l'oeuvre. Citons un exemple(16) de [6] :

Un étudiant est en train de rater un examen oral d'entomologie. Finalement, le professeur lui montre la patte d'un insecte et lui demande d'en identifier le propriétaire. L'étudiant est incapable de le faire, et le professeur le recale. Alors que l'étudiant quitte la salle, le professeur se rend compte qu'il n'a pas noté le nom de l'étudiant. « Quel est ton nom, jeune homme ? », crie-t-il après l'étudiant. L'étudiant revient et tend sa jambe dans la salle et dit : « Devinez-le, professeur ».

En voici un dernier exemple(17) : *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 1 000 ans.*

(16) A student is failing an oral exam in entomology. Finally, the Professor shows him the leg of an insect and asks to identify the owner. The student is unable to do that, and the professor flunks him. As the Student is leaving the room the professor realizes that he did not put down the student's name. « What's your name, young man ? » he shouts after the Student. The student sticks his leg back into the room and says, « You guess it, professor.

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

Les Russes en ont conclu que leurs ancêtres disposaient déjà il y a 1 000 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 2 000 ans.

Les Américains en ont conclu que leurs ancêtres disposaient déjà il y a 2 000 ans d'un réseau de fibre de verre numérique. Et cela, 1 000 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 5 000 ans d'un réseau Wifi ! »

En 2002, Hofstadter [49], grand chantre de l'analogie [50] (Hofstadter et Sander 2013), défend aussi l'idée qu'un mélange de « frames » crée en général de l'incohérence et quelquefois de l'incongruité qui est une bonne source d'humour : « In fact, however, when humans use analogy, most of the time the two spaces are blurred in subtle manners. Such a blurring of distinct situations is termed by some a "frame blend", by others a "blended space". Many blended spaces, though in some sense incoherent, seem completely normal, and often they provide the most natural way of expressing some complicated or subtle idea. Sometimes, however, blended spaces are shockingly incongruous, and this incongruity is the source of humor », adding « humor, or at least a humor-like effect, can arise in an entirely formal domain when blending is done in a sufficiently incoherent manner ».

Les blagues analogiques présentent ainsi des traits généraux communs avec les blagues « logiques » étudiées dans cet article, mais la chute n'y opère par un rétablissement inattendue de la cohérence. À notre connaissance, la formalisation de ce type de blagues n'a pas encore été développée et demanderait donc des études plus approfondies.

6. Conclusion

L'approche proposée met en évidence le rôle clé de la révision des croyances (et de la notion de surprise) dans le mécanisme qui sous-tend les plaisanteries. Elle a été illustrée par quelques exemples. Il est clair qu'une validation de la pertinence de ces idées nécessiterait leur vérification sur un large corpus de blagues, bien que l'on puisse penser que de nombreuses blagues fonctionnent réellement de cette manière. Cette étude montre que la compréhension des blagues peut s'exprimer et se traiter dans le cadre de la représentation des connaissances et du raisonnement sur le changement en intelligence artificielle.

Il arrive que certains articles ou écrits à allure scientifique [87, 102, 120] (Thémiseul de Saint-Hyacinthe 1714, de Unamuno 1946, Pérec 1980) soient des plaisanteries. 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 ¿!

7. Remerciements

Les auteurs remercient les trois relecteurs pour la qualité de leurs rapports et leurs utiles suggestions, et la chatte Aria pour ses interruptions récréatives qui ont ponctué les longues et nombreuses séances de visio lors de l'élaboration de cet article.

Bibliographie

[1] S. Albanie & A. Vedaldi, « Learning grimaces by watching TV », `https://arxiv.org/abs/1610.02255`, 2016.

[2] R. M. Albérès, *Le Comique et l'Ironie*, Thèmes et Parcours Littéraires, Classiques Hachette, 1973.

[3] C. E. Alchourrón, P. Gärdenfors & D. Makinson, « On the logic of theory change: partial meet contraction and revision functions », *The J. of Symbolic Logic* **50** (1985), nº 2, p. 510-530.

[4] S. Attardo, *Linguistic theories of humor*, Mouton, de Gruyter, 1994.

[5] ———, « Formalizing humor theory », in *Proc. 20th Twente Workshop on Language Technology (TWLT20): The April Fools' Day Workshop on Computation Humour, Trento, Apr. 15-16* (O. Stock, C. Strapparava & A. Nijholt, éds.), 2002, p. 1-9.

[6] S. Attardo & V. Raskin, « Script theory revis(it)ed: joke similarity and joke representation model », *Humor-International Journal of Humor Research* **4** (1991), nº 3-4, p. 293-348.

[7] F. Barbieri, F. Ronzano & H. Saggion, « Do we criticise (and laugh) in the same way? Automatic detection of multi-lingual satirical news in Twitter », in *Twenty-Fourth Int. Joint Conf. on Artificial Intelligence*, 2015.

[8] F. Bariaud, *La Genèse de l'Humour chez l'Enfant*, Presses Universitaires De France, Paris, 1983.

[9] G. Bateson, « The role of humor in human communication », in *Cybernetics* (H. von Foerster, éd.), Macy Foundation, New York, 1953.

[10] C. Baudelaire, « De l'essence du rire et généralement du comique dans les arts plastiques », in *Curiosités esthétiques*, Michel Lévy frères, 1868.

[11] J. Beattie, « An essay on laughter, and ludicrous composition », *Essays. William Creech, Edinburgh. Reprinted by Garland, New-York 1971* (1776), p. 414-456.

[12] S. Benferhat, D. Dubois & H. Prade, « Nonmonotonic reasoning, conditional objects and possibility theory », *Artif. Intell.* **92** (1997), nº 1-2, p. 259-276.

[13] ———, « Practical handling of exception-tainted rules and independence information in possibilistic logic », *Appl. Intell.* **9** (1998), nº 2, p. 101-127.

[14] S. Benferhat, D. Dubois, H. Prade & M.-A. Williams, « A practical approach to revising prioritized knowledge bases », *Studia Logica* **70** (2002), nº 1, p. 105-130.

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

[16] K. Binsted & G. Ritchie, « Computational rules for generating punning riddles », **10** (1997), nº 1, p. 25-76.

[17] J.-F. Bonnefon & H. Prade, « Qu'est-ce qui (nous) fait signe ? », in *Dans l'Atelier de l'Art : Expériences Cognitives* (M. Borillo, éd.), Champ Vallon, 2010, p. 186-205.

[18] R. Booth, E. L. Fermé, S. Konieczny & R. P. Pérez, « Credibility-limited improvement operators », in *ECAI*, vol. 263, 2014, p. 123-128.

[19] A. Bourgain, « Rire et angoisse », in *Le Rire à l'épreuve de l'inconscient* (A. Bourgain, C. Chaperot & C. Pisani, éds.), Hermann, 2010, p. 99-119.

[20] A. de. Brahm, *L'Ostensoir des Ironies*, Bibliothèque d'art de la critique, 1899, Réédition partielle par Rumeur des âges (1996), précédé de Le point sur l'ironie par Pierre Schoentjes.

[21] D. Buscaldi & P. Rosso, « Some experiments in humour recognition using the Italian Wikiquote collection », in *Proc. 7th Int. Workshop on Fuzzy Logic and Applications, (WILF'07), Camogli, July 7-10* (F. Masulli, S. Mitra & G. Pasi, éds.), LNCS, vol. 4578, Springer, 2007, p. 464-468.

[22] A. J. Chapman & H. C. Foot (éds.), *Humour and Laughter: Theory, Research and Applications*, John Wiley & Sons, 1976.

[23] C. Darwin, *L'expression des émotions chez l'homme et les animaux*, Reproduction en fac-similé de l'édition de 1890, Complexe, Bruxelles, 1981, 1890, Titre original : The expression of emotions in man and animals, Murray, London, 1872.

[24] Y. Delage, « Sur la nature du comique », *La Revue du Mois* **20** (1919), p. 337-354.

[25] R. Descartes, *Les passions de l'âme*, Henry Le Gras, Paris, 1649.

[26] D. Dubois & H. 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.

[27] ———, « Epistemic entrenchment and possibilistic logic », *Artif. Intell.* **50** (1991), nº 2, p. 223-239.

[28] F. Dupin de Saint-Cyr, « Scenario update applied to causal reasoning », in *Proc. Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'08)*, AAAI Press, septembre 2008, p. 188-197 (english).

[29] F. Dupin de Saint-Cyr & J. Lang, « Belief extrapolation (or how to reason about observations and unpredicted change) », *Artificial Intelligence* **175** (2011) (english), p. 760-790.

[30] F. Dupin de Saint-Cyr & H. Prade, « Peut-on rire en IA ? Révision de croyances et modélisation de plaisanteries », in *Conf. Nat. en Intelligence Artificielle (CNIA'19), Toulouse, Jul. 1-2* (J. Lang, éd.), actes_CNIA@PFIA2019, AfIA, 2019, p. 113-121.

[31] ———, « Jokes and belief revision », in *Proc. 17th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'20), Rhodes, Sept. 12-18* (D. Calvanese, E. Erdem & M. Thielscher, éds.), 2020, p. 336-340.

[32] M. Dynel, « There is method in the humorous speaker's madness: Humour and Grice's model », *Lodz Papers in Pragmatics* **4** (2008), nº 1, p. 159-185.

[33] R. Epstein & V. R. Joker, « A threshold theory of the humor response », *The Behavior Analyst* **30** (2007), nº 1, p. 49-58.

[34] R. Fagin & J. Y. Halpern, « Reasoning about knowledge and probability », *J. of the ACM (JACM)* **41** (1994), nº 2, p. 340-367.

[35] F. R. Fleet, *A Theory of Wit and Humour*, Remington and Company, 1890.

[36] S. 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, 1961*, 1ère publ. : Int. J. of Psychoanalysis, 9, 1-6, 1928, 1928, p. 161-166.

[37] ———, *The Joke and its Relation to the Unconscious*, Penguin, 2003, première édition allemande en 1905. Traduction française : Le mot d'esprit et sa relation à l'inconscient, Gallimard 1992.

[38] P. Gärdenfors, *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, MIT Press, 1988, 2nd ed., College Publications, 2008.

[39] ———, « Belief revision and nonmonotonic logic: Two sides of the same coin\string ? », in *Proceedings of the 9th European Conference on Artificial Intelligence (ECAI'90), Stockholm*, 1990, p. 768-773.

[40] J. H. Goldstein & P. E. McGhee (éds.), *The Psychology of Humor. Theoretical Perspectives and Empirical Issues*, Academic Press, 1972.

[41] J. C. Gregory, *The nature of laughter*, Harcourt, Brace, 1924.

[42] P. Grice, « Logic and conversation », in *Syntax and Semantics, vol 3* (P. Cole & J. Morgan, éds.), Academic Press, New York, 1975.

[43] A. Grove, « Two modellings for theory change », *Journal of Philosophical Logic* **17** (1988), nº 2, p. 157-170.

[44] W. N. Guthrie, « A theory of the comic », *International Quarterly* **7** (1903), p. 254-264.

[45] S. O. Hansson, E. L. Fermé, J. Cantwell & M. A. Falappa, « Credibility limited revision », *The Journal of Symbolic Logic* **66** (2001), nº 4, p. 1581-1596.

[46] M. Hasan, S. Lee, W. Rahman, A. Zadeh, R. Mihalcea, L. Morency & E. Hoque, « Humor knowledge enriched transformer for understanding multimodal humor », in *Proc. 35th AAAI Conf. on Artificial Intelligence (AAAI'2021), February 2-9*, 2021, p. 12972-12980.

[47] C. Hempelmann, V. Raskin & K. E. Triezenberg, « Computer, tell me a joke... but please make it funny: Computational humor with ontological semantics », in *Proc. 19th Int. Florida Artificial Intelligence Research Society Conf., FLAIRS'06* (G. Sutcliffe & R. Goebel, éds.), AAAI Press, 2006, p. 746-751.

[48] T. Hobbes, *Leviathan*, Crooke, London Press, 1651.

[49] D. Hofstadter, « Frame blends, analogies, and humor – Abstract », in *Proc. 20th Twente Workshop on Language Technology (TWLT20): The April Fools' Day Workshop on Computation Humour, Trento, Apr. 15-16* (O. Stock, C. Strapparava & A. Nijholt, éds.), 2002, p. 29.

[50] D. Hofstadter & E. Sander, *Surfaces and Essences: Analogy as the Fuel and Fire of Thinking*, Basic Books, 2013, Version française : L'Analogie, Cœur de la Pensée - Odile Jacob, Paris, 2013.

[51] H. L. Hollingworth, « Experimental studies in judgment: Judgments of the comic », *Psychological Review* **18** (1911), nº 2, p. 132-156.

[52] L. Hunt, *Wit and Humour: Selected from the English Poets*, Smith, Elder, 1848.

[53] M. M. Hurley, D. C. Dennett, R. B. Adams Jr & R. B. Adams, *Inside Jokes: Using Humor to Reverse-Engineer the Mind*, MIT press, 2011.

[54] V. Jankélévitch, *L'Ironie*, Flammarion, 1964.

[55] L. Joubert, *Traité du Ris Contenant son Essance, ses Causes et Mervelheus Effais, Curieusemant Recerchés, Raisonnés, et Observés*, Nicolas Chesneau, Paris, 1579, disponible sur Gallica. Traduction anglaise : Treatise on Laughter. Translation from French of an Essay on the Nature and Character of Human Laughter, University Alabama Press, 2016.

[56] J. Kagan, « Discrepancy, temperament, and infant distress », *The origins of fear* (1974), p. 229-248.

[57] E. Kant, *Critique de la Faculté de Juger*, Garnier-Flammarion, 1790.

[58] J. T. Kao, R. Levy & N. D. Goodman, « A computational model of linguistic humor in puns », **40** (2016), nº 5, p. 1270-1285.

[59] J. Karoui, F. Benamara & V. Moriceau, *Détection automatique de l'ironie : Application à la fouille d'opinion dans les microblogs et les médias sociaux*, ISTE Group, 2019.

[60] H. Katsuno & A. O. Mendelzon, « On the difference between updating a knowledge base and revising it », *Belief revision* **29** (1991), p. 183.

[61] ———, « Propositional knowledge base revision and minimal change », *Artificial Intelligence* **52** (1991), nº 3, p. 263-294.

[62] P. Keith-Spiegel, « Early conceptions of humor: Varieties and issues », in *The Psychology of Humor: Theoretical Perspectives and Empirical Issues* (J. H. Goldstein, éd.), Academic Press New York, 1972, p. 3-39.

[63] É. Klein & J. Perry-Salkow, *Anagrammes Renversantes : ou Le Sens Caché du Monde*, Flammarion, 2011.

[64] L. W. Kline, « The psychology of humor », *The American Journal of Psychology* (1907), p. 421-441.

[65] S. Konieczny & R. Pino Pérez, « A framework for iterated revision », *J. Applied Non-Class. Logics* **10** (2000), p. 339-367.

[66] M. Latour, *Le Problème du Rire et du Réél. Quatre Essais*, Presses Universitaires De France, Paris, 1956.

[67] D. Lehmann, « Belief revision, revised », in *Proceedings of the 14th international joint conference on Artificial intelligence-Volume 2*, 1995, p. 1534-1540.

[68] G. Lessard & M. Levison, « Computational modelling of linguistic humour: *Tom Swifties* », in *Proc. 19th Int. Conf. of the Association For Literary An d Linguistic Computing and the 12th Int. Conf. on Computing and Humanities (ALLC-ACH '92), Oxford*, 1992.

[69] G. Lessard, M. Levison & C. Venour, « Cleverness versus funniness », in *Proc. of the 20th Twente Workshop on Language Technology (TWLT20): The April Fools' Day Workshop on Computation Humour, Trento, Apr. 15-16* (O. Stock, C. Strapparava & A. Nijholt, éds.), 2002, p. 137-145.

[70] M. Letourneux & A. Vaillant (éds.), *L'Empire du Rire. XIXe - XXIe siècle*, CNRS éditions, 2021.

[71] E. Lorini & C. Castelfranchi, « The cognitive structure of surprise: looking for basic principles », *Topoi* **26** (2007), nº 1, p. 133-149.

[72] A. M. Ludovici (éd.), *The Secret of Laughter*, Constable & Co Ltd,London, 1932.
[73] N. R. F. Maier, « A Gestalt theory of humour », *British Journal of Psychology* **23** (1932), nº 1, p. 69-74.
[74] R. A. Martin & T. E. Ford (éds.), *The Psychology of Humor An Integrative Approach*, 2nd éd., Academic Press, Elsevier, 2018.
[75] J. McKay, « Generation of idiom-based witticisms to aid second language learning », in *Proc.of the 20th Twente Workshop on Language Technology (TWLT20): The April Fools' Day Workshop on Computation Humour, Trento, Apr. 15-16* (O. Stock, C. Strapparava & A. Nijholt, éds.), 2002, p. 77-87.
[76] S. Melchior-Bonnet, *Le Rire des Femmes*, PUF, 2021.
[77] F. Mele, « Real and fictional ridicule. A cognitive approach for models of humour », in *Proc.of the 20th Twente Workshop on Language Technology (TWLT20): The April Fools' Day Workshop on Computation Humour, Trento, Apr. 15-16* (O. Stock, C. Strapparava & A. Nijholt, éds.), 2002, p. 89-99.
[78] R. Mihalcea, « The multidisciplinary facets of research on humour », in *Proc. 7th Int. Workshop on Fuzzy Logic and Applications, (WILF'07), Camogli, July 7-10* (F. Masulli, S. Mitra & G. Pasi, éds.), LNCS, vol. 4578, Springer, 2007, p. 412-421.
[79] R. Mihalcea & C. Strapparava, « Learning to laugh (automatically): Computational models for humor Recognition », *Comput. Intell.* **22** (2006), nº 2, p. 126-142.
[80] R. Mihalcea, C. Strapparava & S. G. Pulman, « Computational models for incongruity detection in humour », in *Proc. 11th Int. Conf. Computational Linguistics and Intelligent Text Processing (CICLing'10), Iasi, March 21-27* (A. F. Gelbukh, éd.), LNCS, vol. 6008, Springer, 2010, p. 364-374.
[81] M. Minsky, « Jokes and the logic of the cognitive unconscious », in *Cognitive Constraints on Communication - Representations and Processes* (L. Vaina & J. Hintikka, éds.), D. Reidel Publ. Comp., 1984, p. 175-200.
[82] D. H. Monro, *Argument of Laughter*, Melbourne University Press, 1951.
[83] H. Nègre, *Dictionnaire des Histoires Drôles*, Fayard, 1970.
[84] R. Niewiadomski, K. Prepin, E. Bevacqua, M. Ochs & C. Pelachaud, « Towards a smiling ECA: studies on mimicry, timing and types of smiles », in *Proceedings of the 2nd International Workshop on Social Signal Processing*, 2010, p. 65-70.
[85] M. Ochs, C. Pelachaud & G. Mckeown, « A User Perception–Based Approach to Create Smiling Embodied Conversational Agents », *ACM Transactions on Interactive Intelligent Systems (TiiS)* **7** (2017), nº 1, p. 1-33.
[86] M. Pagnucco, *The Role of Abductive Reasoning within the Process of Belief Revision*, Univ. Sydney Dissertation, 1996.
[87] G. Pérec, *Cantatrix Sopranica L. et Autres Ecrits Scientifiques*, Éditions du Seuil, coll. « Librairie du XXe siècle », 1980 (réédité en 1995).
[88] J. Perry-Salkow & L. Castelain, *Anagrammes dans le Boudoir*, Actes Sud, 2020.
[89] H. Prade, « Monde(s) et représentation(s). Quelques propos sur les rapports entre art et science », in *Approches Cognitives de la Création Artistique* (M. Borillo, éd.), Mardaga, 2005, p. 49-56.
[90] F. Rabelais, *La Vie Très Horrificque du Grand Gargantua, Père de Pantagruel, Jadis Composée par M. Alcofribas Abstracteur de Quintessence*, La Sirène (réédition 1919), 1534.
[91] P.-Y. 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), p. 75-94.
[92] V. S. Ramachandran, « The artful brain », in *The Internet and the University: Forum 2004*, Fourth Estate, 2004, p. 169-198.
[93] J. D. Raskin & S. Attardo, « Non-literalness and non-bona-fide in language: An approach to formal and computational treatments of humor », *Pragmatics and Cognition* **2** (1994), nº 1, p. 31-69.
[94] V. Raskin, *Semantic Mechanisms of Humor*, Dordrecht, Reidel, 1984.
[95] J. M. Raulin, *Le rire et les exhilarants. Etude anatomique, psycho-physiologique et pathologique*, J. B. Baillière, Paris, 1900.
[96] G. Ritchie, « Developing the incongruity-resolution theory », in *Proc. of the AISB Symposium on Creative Language: Stories and Humour, Edinburgh*, 1999, p. 78-85.

[97] ———, « The structure of forced reinterpretation jokes », in *Proc. 20th Twente Workshop on Language Technology (TWLT20): The April Fools' Day Workshop on Computation Humour, Trento, Apr. 15-16* (O. Stock, C. Strapparava & A. Nijholt, éds.), 2002, p. 47-56.

[98] ———, « Logic and reasoning in jokes », *The Europ. J. of Humour Research* **2** (2014), nº 1, p. 50-60.

[99] ———, *The Comprehension of Jokes: A Cognitive Science Framework*, CRC Press, 2018.

[100] J. E. Roeckelein, *The Psychology of Humor: A Reference Guide and Annotated Bibliography*, Greenwood Press, 2002.

[101] W. Ruch, S. Attardo & V. Raskin, « Toward an empirical verification of the general theory of verbal humor », *Humor: International Journal of Humor Research* **6** (1993), nº 2, p. 123-136.

[102] T. de Saint-Hyacinthe, « Le Chef d'Œuvre d'un Inconnu », 1714, Réédition Aubanel (1965).

[103] P. Schiller, « A configurational theory of puzzles and jokes », *The Journal of General Psychology* **18** (1938), nº 2, p. 217-234.

[104] A. Schopenhauer, « Le Monde comme volonté et comme représentation », 1819.

[105] G. L. S. Shackle, *Decision, Order and Time in Human Affairs* (2nd edition), Cambridge University Press, UK, 1961.

[106] D. Shahaf, E. Horvitz & R. Mankoff, « Inside jokes: Identifying humorous cartoon captions », in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2015, p. 1065-1074.

[107] Y. Shoham & S. B. Cousins, « Logics of mental attitudes in AI. A very preliminary survey », in *Foundations of Knowledge Representation and Reasoning* (G. Lakemeyer & B. Nebel, éds.), LNCS, vol. 810, Springer, Berlin, Heidelberg, 1994, p. 296-309.

[108] T. R. Shultz, « A cognitive-developmental analysis of humour », in *Humour and Laughter: Theory, Research and Applications* (A. J. Chapman & H. C. Foot, éds.), John Wiley & Sons, 1976, p. 11-36.

[109] D. Sjöbergh & K. Araki, « Recognizing humor without recognizing meaning », in *Proc. 7th Int. Workshop on Fuzzy Logic and Applications, (WILF'07), Camogli, July 7-10* (F. Masulli, S. Mitra & G. Pasi, éds.), LNCS, vol. 4578, Springer, 2007, p. 469-476.

[110] E. Smadja, *Le Rire*, Collection Que Sais-Je ?, Presses Universitaires de France, Paris, 1993.

[111] L. Sombé (collectif), « A glance at revision and updating in knowledge bases », *International J. of Intelligent Systems* **9** (1994), nº 1, p. 1-27.

[112] H. Spencer, « The physiology of laughter », *Macmillan's Magazine* **1** (1860), p. 395-402.

[113] O. Stock & C. Strapparava, « Getting serious about the development of computational humor », in *Proc. IJCAI'03*, vol. 3, 2003, p. 59-64.

[114] J. Sully, *Essay on laughter*, Longmans, Green, New York, 1902.

[115] J. M. Suls, « A two-stage model for the appreciation of jokes and cartoons: An information-processing analysis », *The psychology of humor: Theoretical perspectives and empirical issues* **1** (1972), p. 81-100.

[116] M. Tahon & L. Devillers, « Laughter detection for on-line human-robot interaction », in *Interdisciplinary Workshop on Laughter and Non-verbal Vocalisations in Speech*, 2015.

[117] G. Tibballs (éd.), *The Mammoth Book of Jokes*, Constable & Robinson, London, 2000.

[118] H. W. Tinholt, « Computational Humour. Utilizing cross-reference ambiguity for conversational jokes », Master's thesis, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science, 2007.

[119] H. W. Tinholt & A. Nijholt, « Computational humour: Utilizing cross-reference ambiguity for conversational jokes », in *Proc. 7th Int. Workshop on Fuzzy Logic and Applications, (WILF'07), Camogli, July 7-10* (F. Masulli, S. Mitra & G. Pasi, éds.), LNCS, vol. 4578, Springer, 2007, p. 477-483.

[120] M. de. Unamuno, *La Cocotologie. Notes pour un Traité*, Editions Self, Paris, 1946.

[121] P. Vrticka, J. M. Black & A. L. Reiss, « The neural basis of humour processing », *Nature Reviews Neuroscience* **14** (2013), nº 12, p. 860-868.

[122] M.-A. Williams, « Aesthetics and the explication of surprise », *Languages of Design. Formalisms for word, image and sound* **3** (1996), p. 145-157.

[123] M. Winslett, « Reasoning about action using a possible models approach », in *Proc. AAAI'88*, 1988, p. 89-93.

[124] M. Yaari, *Ironie Paradoxale et Ironie Poétique*, Summa Publications, Birmingham, Alabama, 1988.

[125] S. Zeki, *Inner Vision. An Exploration of Art and the Brain*, Oxford University Press, 1999.
[126] E. Zigler, J. Levine & L. Gould, « Cognitive challenge as a factor in children's humor appreciation », *Journal of Personality and Social Psychology* **6** (1967), nº 3, p. 332-336.

Abstract. — There are some human cognitive activities that have remained little studied in artificial intelligence : understanding laughter is an example. This exploratory paper tries to identify the mechanisms at work in quips and narrative logical jokes. It seems that in many cases a belief revision mechanism is operating, leading to an unexpected conclusion, through the punchline of the jest. The jokes with analogical functioning are also evoked. The article also includes a substantial bibliographical study covering ancient or modern works on laughter (illustrated by some puns).

Keywords. — Humor, belief revision, surprise, inconsistency, incongruity.

Manuscrit reçu le 11 juin 2021, révisé le 10 août 2021, accepté le 11 septembre 2021.