

# Computational Humor

## Eine Übersicht über maschinelles Humorverständnis

Timo Hardebusch

TU Dortmund

`timo.hardebusch@tu-dortmund.de`

**Abstract.** Humor is an essential part of human existence and yet our understanding of it is still incomplete. Fooling around is a common element of human communication, but people no longer only interact with other people. A growing number of machine assistants and companions are becoming part of everyday life. In this seminar paper an overview of the field of Computational Humor is given, i. e. the understanding and generation of humor by computers. To this end, the theoretical foundations of humor are first examined from the perspective of various disciplines such as psychology and neuroscience. This is followed by an explanation of the Incongruity Resolution Model, which is a predominant theory in the field of humor research. It provides the basis for the subsequently presented approaches to make humor comprehensible for computers, such as the generation of puns, funny acronyms and clipart pictures or the recognition of funny one-liners. While the results of the presented methods are not at a human-like level, they provide important fundamentals for future development in this area and humour research itself.

## 1 Einleitung

*Verloren ist, wer den Humor verlor.*  
(Otto Julius Bierbaum, 1906)

Die Fähigkeit zu verlieren, über kleine und größere Missgeschicke zu lachen und dem Schicksal mit einer gesunden Portion Ironie gegenüberzustehen – diese Vorstellung dürfte den meisten Menschen Unbehagen bereiten. Kalt und *roboterhaft* erscheint ein Leben ohne Heiterkeit. Und in der Tat: Selbst Hollywood tut sich schwer damit, futuristischen Maschinenwesen einen Sinn für Humor zuzuschreiben. So wird etwa der Androide *Data* in der Serie *Raumschiff Enterprise: Das nächste Jahrhundert* neben seinem enormen Wissensschatz vor allem durch eine charakteristische Abwesenheit von jeglichem Humorverständnis porträtiert. Erst als ihm im Serienverlauf ein *Emotionschip* eingesetzt wird, eröffnet sich ihm – mitunter zum Leidwesen seiner Mannschaftskameraden – die Welt der Komik und Kalauer.

Obgleich von Menschen nahezu ununterscheidbare Androiden noch den Anschein einer fernen Zukunftsvision haben, wird der Alltag zunehmend durch die

Interaktion mit intelligenten Rechnersystemen durchsetzt. Mit einer Sprachsteuerung versehene persönliche Agenten, Schnittstellen von *Smart-Home*-Systemen, der durch einen Bot verwaltete *First-Level-Support*: All diese Systeme könnten durch das Verständnis und die dosierte eigene Anwendung von Humor ein natürlicheres Kommunikationserlebnis für ihre menschlichen Benutzer bieten.

Doch wie realistisch ist der sinnbildliche humorstiftende Emotionschip? In dieser Arbeit wird ein Überblick über ausgewählte Versuche gegeben, Rechnersystemen das Erkennen und Benutzen unterschiedlicher Ausprägungen von Humor beizubringen: *Computational Humor*. Dazu wird in Abschnitt 2 zunächst eine Einordnung *menschlichen* Humors aus der Sicht einschlägiger Wissenschaftszweige skizziert. In Abschnitt 3 folgt eine Beschreibung von Übertragungsansätzen dieser Erkenntnisse in die Welt der Rechnersysteme. In Abschnitt 4 wird abschließend eine Zusammenfassung der vorgestellten Verfahren und ein kurzer Ausblick auf offene Probleme und weiterführende Fragestellungen gegeben.

## 2 Humor – Interdisziplinäre Perspektiven

Trotz der alltäglichen Konfrontation und des üblicherweise intuitiven Verständnisses von Humor durch den Menschen, gestaltet sich eine umfängliche Erfassung der Thematik nicht trivial. In diesem Abschnitt werden orientiert an der Literatur [1,2] Lösungsansätze zu den Fragen skizziert, *warum* und *worüber* Menschen lachen.

### 2.1 Versuch einer Einordnung

Eine grundlegende Schwierigkeit stellt sich bereits bei der Begriffsdefinition dar: Im Allgemeinen wird Humor als Synonym oder Hyperonym für verschiedene Verhaltensweisen oder Ausdrucksformen genutzt, einschließlich, aber nicht beschränkt auf Scherze, Wortspiele, Ironie, Satire, Hohn, Comic und Tragikomik. Es besteht jedoch kein Konsens darüber, welche dieser Elemente zwingend oder optional unter der Definition subsumiert werden sollten. Dass eine durch die angeführten Beispiele angedeutete Heterogenität sich dabei nicht auf die begriffliche Ebene beschränkt, bekräftigen Forschungsergebnisse aus dem Bereich der Neurowissenschaften. Mittels bildgebender Verfahren lassen sich Unterschiede in der Verarbeitung verschiedener Ausdrucksformen von Humor im menschlichen Gehirn erkennen. So kann beispielsweise zwischen unwillkürlichem, also reflexartigem, und bewusstem Lachen anhand der beteiligten Gehirnareale differenziert werden [3]; auch unterliegen Wortspiele einer anderen Verarbeitung als Witze, die auf semantischem Verständnis beruhen [4].

Diese strukturellen Unterschiede stützen Theorien der Evolutionspsychologie, nach denen das unwillkürliche Lachen, das in evolutionär älteren Hirnarealen ausgelöst wird, eine ältere Verhaltensweise als das bewusst gesteuerten Lachen ist. Tatsächlich lassen sich auch bei anderen Spezies, darunter bestimmte Gattungen Affen und Hunde, charakteristische Laute und Umgangsformen beobachten, die mit spielerischen und freudigen Interaktionen einhergehen. Auch beim Menschen

wird davon ausgegangen, dass Lachen und Humor eine sozial förderliche Wirkung haben: So können Spannungen zwischen Individuen und Gruppen abgebaut werden und der Nachwuchs wird in die Lage versetzt, wichtige Verhaltensweisen in einem spielerischen Kontext zu erproben, beispielsweise der Kampf mit Altersgenossen oder Eltern.

Auch wenn diese Ausführungen eine gewisse Universalität des redensartlichen Sinnes für Humor nahelegen, ist dessen Ausprägung individuell unterschiedlich. Neben persönlichen Eigenschaften (u. a. Intelligenz, Bildungsgrad, Alter) sind auch Umgebungsfaktoren wie die regionale Kultur von Bedeutung. Zusammenfassend lässt sich festhalten, dass Humor ein heterogenes Konzept ist, dessen Hintergründe weiterhin Gegenstand der Forschung verschiedener Disziplinen sind und im Kontext dieser Ausarbeitung dementsprechend nur oberflächlich behandelt werden kann.

## 2.2 Incongruity-Resolution-Modell

Trotz der im vorangegangenen Abschnitt beschriebenen Uneinheitlichkeit gibt es Theorien, die gemeinsame Strukturen verschiedener Ausdrucksformen von Humor feststellen. Ein vorherrschendes Konzept ist das *Incongruity-Resolution*-Modell, das nachfolgend erläutert wird.

Nach diesem Modell wird beim Adressaten einer humorvollen Schilderung zunächst eine gewisse Erwartungshaltung aufgebaut; dies geschieht meist durch das implizite Aufdrängen einer offensichtlichen Interpretation des Beschriebenen. Anschließend folgt eine Pointe, die im Widerspruch zu dieser Vorstellung steht (*Incongruity*). Ein solcher kann syntaktischer Natur sein (bspw. ein unpassender Kasus oder Genus) oder durch unerwartete Semantik entstehen und durch Verwendung kontroverser Themen wie Sexualität oder Obszönität zusätzlich verstärkt werden. Dieser Bruch führt zu einer Reevaluation durch den Empfänger. Unter Kenntnis der Pointe eröffnet sich eine alternative Deutung des Vorangegangenen, was – vorausgesetzt, der Humor wird wie intendiert verstanden – zu einer Auflösung des Konfliktes führt (*Resolution*) und mit dieser zu einer Erleichterung und Erheiterung. [5]

Zum besseren Verständnis soll beispielhaft der folgende Witz betrachtet werden: *Treffen sich zwei Jäger im Wald. Beide tot.* In der Exposition wird eine Vorstellung nahegelegt, dass es sich bei der beschriebenen Situation um eine Begegnung oder Verabredung („treffen sich“) zweier Personen handelt. Die Pointe („beide tot“) passt zunächst nicht zu dieser Interpretation; es entsteht eine Diskrepanz zwischen dem erwarteten Fortschreiten der Erzählung und dem überraschenden Ableben der Protagonisten. Eine Reevaluation offenbart die zunächst verborgene Deutung: Die Jäger haben sich durch Gebrauch ihrer Schusswaffen anscheinend gegenseitig mit einem Projektil todbringend *getroffen*. Erst dieser Auflösungsprozess sorgt für den erfolgreichen Witz.

Einschränkend ist anzumerken, dass das *Incongruity-Resolution*-Modell nicht jeglichen Aspekt von Humor vollständig beschreiben kann. Ungeklärt bleibt in diesem Rahmenwerk beispielsweise, warum nicht jede Auflösung von Widersprüchlichkeiten zu humorvollen Reaktionen führt oder ein Witz auch nach

mehrfachem Hören, also bei bereits bekannter „korrekter“ Lesart, noch als lustig wahrgenommen werden kann. Weiterhin fehlt eine einheitliche Auffassung, *wie genau* der Resolutionsprozess zu verstehen ist, ob dieser im Einklang mit der Wortherkunft eine tatsächliche *Auflösung* der Inkongruenz bedingt oder lediglich eine Art *kongruente Inkongruenz* herstellt. [5]

### 3 Computational Humor

In diesem Abschnitt werden nach einer kurzen allgemeinen Problemanalyse Ansätze beschrieben, wie Rechnersystemen der Umgang mit verschiedenen Formen von Humor ermöglicht werden kann. Dabei kann angesichts des begrenzten Umfangs der Ausarbeitung lediglich ein Überblick gegeben werden; für vertiefende Ausführungen sei auf die jeweils angeführte Literatur verwiesen.

#### 3.1 Inhärente Herausforderungen

Wie im vorangegangenen Abschnitt dargelegt, hat menschlicher Humor mannigfaltige Ausprägungsformen, deren Verständnis und Bewertung sowohl inter- als auch intrapersonal unterschiedlich ausfällt. Diese Komplexität deutet die Schwierigkeiten an, die eine umfassende Lösung für maschinelles Humorverständnis innehat; es wird von einem *KI-vollständigen* (*AI-complete*) [6] Problem ausgegangen [7], dessen Lösung die Fähigkeiten einer *Artificial General Intelligence* oder *starken KI* erfordert.

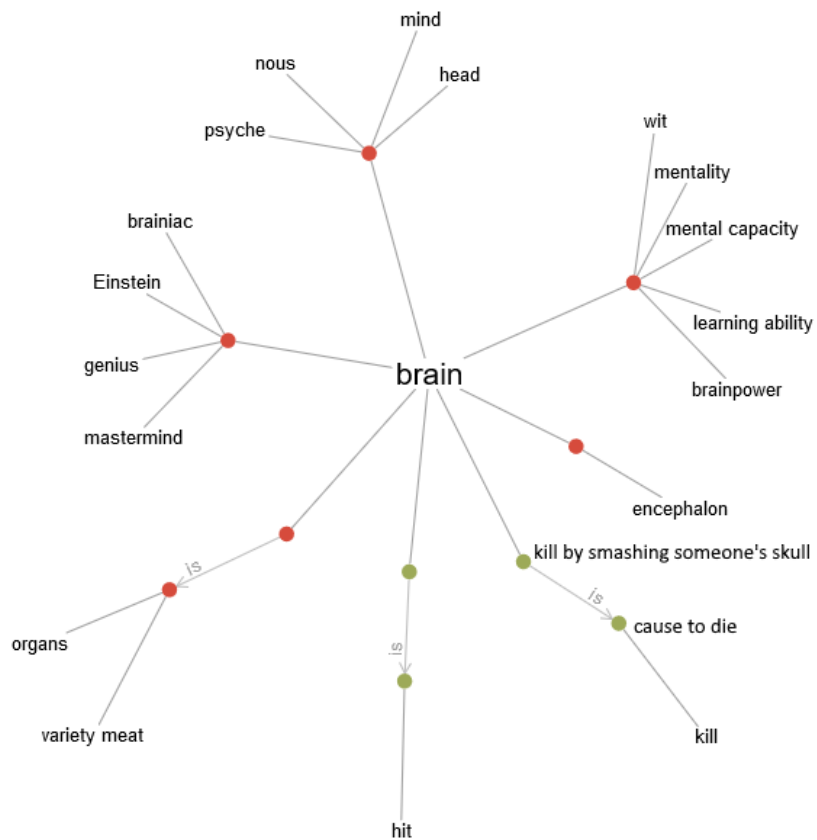
Auch ein Ansatz über das *Incongruity-Resolution*-Modell, das auf der strukturellen Ebene von humorvollen Ausdrucksformen basiert, erfordert umfangreiches semantisches, domänenübergreifendes Wissen über die Welt. Eine lexikalische Definition von Begriffen ist nicht ausreichend, um alle Ausprägungen von Inkongruenz zu erfassen; viele, insbesondere subtilere Formen von Humor setzen ein Verständnis davon voraus, wie Objekte sich *üblicherweise* verhalten, wie sie zueinander und mit ihrer Umwelt in Verbindung stehen. (Gesunde) Menschen können dazu sowohl auf das semantische Gedächtnis zurückgreifen, das allgemeines Weltwissen enthält, als auch auf das episodische Gedächtnis, das persönliche Erfahrungen beinhaltet [8].

Einen Lösungsansatz, der bei einigen der nachfolgend vorgestellten Konzepte genutzt wird, stellt das lexikalisch-semantische Netz *WordNet* [9] dar, das für die automatische Verarbeitung durch Rechnersysteme ausgelegt ist. In diesem werden Begriffe anhand semantischer Relationen (bspw. synonym, antonym, hyperonym, hyponym) einander zugeordnet. Abbildung 1 zeigt dies an einem Beispiel.

Einschränkend ist anzumerken, dass WordNet nur die englische Sprache berücksichtigt. Übersetzungen und alternative Projekte unterschiedlichen Umfangs sind jedoch Gegenstand anhaltender Arbeiten [10,11].

#### 3.2 JAPE und STANDUP

**JAPE.** *JAPE (Joke Analysis and Production Engine)* ist ein in den 1990er Jahren entwickeltes Programm, das kurze Frage-Antwort-Rätsel, die auf Wortspielen



**Abbildung 1.** Visualisierung des WordNet-Eintrags für *brain* (Gehirn). Es sind sowohl Nomen zugeordnet, etwa als Synonym für die geistigen Fähigkeiten (*mental capacity*) oder als Bezeichnung für ein Genie (*genius*), als auch Verben wie töten (*kill*) durch einen letalen Schlag auf den Schädel.

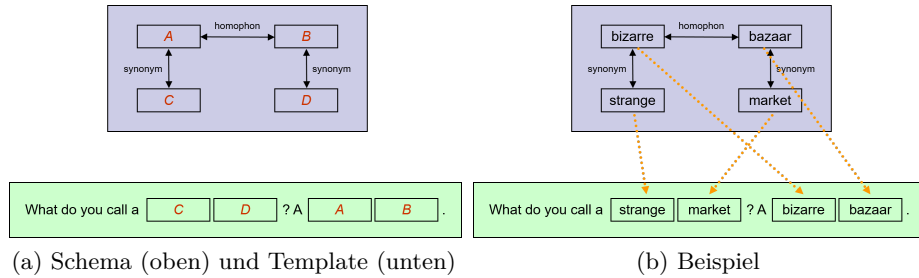
basieren, generiert [12]. Beispiele für solche von JAPE erstellten Scherzfragen sind:

*Beispiel 1. What is the difference between a pretty glove and a silent cat?  
One is a cute mitten, the other is a mute kitten.*

*Beispiel 2. What kind of emotion has bits? A love byte.<sup>1</sup>*

Das System benötigt für seine Funktion umfangreiche menschliche Vorarbeit: Es müssen manuell in einer *Prolog*-ähnlichen Notation *Schemata* und *Templates* erstellt werden, die JAPE als Grundlage für die Suche nach passenden Begriffspaaren in WordNet verwendet. Durch diese werden Vorgaben getroffen, die die

<sup>1</sup>Hier könnte gar von *Computational Computer Humor* gesprochen werden.



**Abbildung 2.** Visualisierung der Anwendung von Schemata und Templates bei JAPE. Für die händisch festgelegten Rahmenbedingungen werden Begriffe, die diese Eigenschaften erfüllen, in WordNet gesucht und eingesetzt. (Abbildung modifiziert nach [13])

Struktur der generierten Wortspiele festlegt. Abbildung 2 zeigt dies an einem Beispiel. In einer späteren Version von JAPE werden zusätzlich *Beschreibungen* eingeführt, die einen Schritt zwischen Schema und Template bilden. In diesen werden minimal notwendige Beschreibungen zu den Begriffen abstrahiert, die die Schemata erfüllen, um auf diese Weise sinnvollere Fragen zu generieren. Die Charakterisierung einer Zitrone als gelbe Zitrusfrucht ist beispielsweise besser geeignet als die eines gelben Objektes, das Kerne enthält.

Bei der Gestaltung der Schemata und Templates kann das *Incongruity-Resolution*-Modell herangezogen werden. Im Beispiel 1 wird die Inkongruenz bereits in der Frage erzeugt, indem zwei Begriffe aus unterschiedlichen Domänen verglichen werden (Handschuh vs. Katze), was erst durch die Auflösung in der Antwort einen gewissen, wenn auch nur phonetischen, Sinn bekommt. Im Beispiel 2 wird eine ähnliche Strategie genutzt wie im Jäger-Beispiel in Abschnitt 2.2: Der Begriff *love byte* (Verballhornung von *love bite*, Knutschfleck) sorgt zunächst für Unklarheit, welche unter Reevaluation der Mehrdeutigkeit von *bits* (Häppchen/Gebisse vs. Bits als *binary digits*, Binärziffern) aufgelöst wird.

Problematisch an dieser Methode der Humorgenerierung ist neben der Abhängigkeit von engen menschlichen Vorgaben eine mäßige Qualität der Ergebnisse. Während bei einer Auswertung mit 120 Kindern im Grundschulalter die von JAPE generierten Wortspiele durchschnittlich als lustiger eingestuft wurden als nicht witzige Frage-Antwort-Paare, die als Kontrolle dienten, wurden sie als signifikant weniger witzig als von Menschenhand erstellte Scherzfragen bewertet. [14] Dabei ist allerdings zu berücksichtigen, dass WordNet zur Entwicklungszeit des JAPE-Projekts weniger Funktionen (bspw. keine Verben, weniger semantische Relationen) zur Verfügung gestellt hat, sodass einzelne Schemata und Templates nur unzureichend umsetzbar waren.

JAPE bietet weiterhin nur eingeschränkte Konfigurationsmöglichkeiten, um aus den statischen Vorgaben abwechslungsreichere Wortspiele zu generieren. Eine graphische Benutzerschnittstelle ist ebenfalls nicht implementiert.

**STANDUP.** Eine Weiterentwicklung dieses Konzepts stellt *STANDUP* (*System To Augment Non-speakers' Dialogue Using Puns*) dar. [15] Das Ziel dieses Projektes ist, eine Lernumgebung für Kinder mit kommunikativen und motorischen Schwächen zu schaffen, um deren Sprachentwicklung zu fördern. Durch einen interaktiven Prozess, der humorvolle Frage-Antwort-Rätsel nach dem Vorbild von JAPE generiert, sollen die Kinder ermutigt werden, Sprache spielerisch zu erkunden, indem sie sich mit Elementen wie Mehrdeutigkeiten und Betonung von Ausdrücken auseinandersetzen.

Während der grundlegende Ablauf des Generators mit Schemata, Beschreibungen und Templates in STANDUP beibehalten wird, sorgen Modifikationen in den Auswahlprozessen passender Begriffe für verfeinerte und vielfältigere Ergebnisse. Es werden zwischenzeitlich hinzugekommene Funktionen von WordNet genutzt, beispielsweise die Verwendung von Verben, sowie eine Erweiterung der Homophonie von gleich klingenden zu annähernd ähnlich ausgesprochenen Worten (z. B. *rave* ['rev] und *reef* ['ri:f]). Um einen interaktiven Prozess zu ermöglichen, können Themenfelder eingegrenzt werden. Außerdem ist es möglich, im Hinblick auf die Zielgruppe anstößige oder in der Alltagssprache selten verwendete Formulierungen auszuschließen.

Im Gegensatz zu JAPE verfügt STANDUP über eine graphische Benutzerschnittstelle, die den Anwender durch die Schritte des Generierungsprozesses führt und durch Symbole beim Verständnis der Witze hilft; ein Ausschnitt davon ist in Abbildung 3 zu sehen. Ein integriertes *Text-to-Speech*-System ermöglicht das Vorlesen der Rätsel, um die Aussprache der Wortspiele zu üben.

### 3.3 HAHAAcronym

Das Ziel des *HAHAAcronym*-Projekts (*Humorous Agents for Humorous Acronyms*) ist die humorvolle Abwandlung bestehender oder die Generierung neuer lustiger Akronyme, d. h. Abkürzungen, die aus den Anfangsbuchstaben mehrerer Worte bestehen [16]. Ein Beispiel ist *MIT* für das *Massachusetts Institute of Technology*.

Als Ressource für Begriffe wird auch hier WordNet eingesetzt, zusammen mit der Erweiterung *WordNet Domains*, die Wörter Domänen zuordnet (bspw. *nurse* der Domäne *medicine*) [17]. In späteren Versionen wird außerdem *WordNet Affect* eingebunden, das WordNet-Einträge mit emotionalen Attributen versieht (bspw. *ghost* mit der Eigenschaft *frightening*) [18]. Die grundlegende Idee ist auch bei diesem Ansatz, Inkongruenz zu erzeugen. Dies wird erreicht, indem Begriffe aus gegensätzlichen Domänen oder emotionalen Charakteristiken miteinander verbunden werden (bspw. *science* vs. *religion*, *frightening* vs. *comforting*). Zur Modifikation eines übergebenen Akronyms wird dieses zunächst unter einer Akronymgrammatik verarbeitet; dies geschieht mit einem *Augmented-Transition-Network*-Parser. Der dabei identifizierte Kopf der Nominalphrase (d. h. üblicherweise ein Substantiv oder Pronomen) wird beibehalten, die übrigen Worte werden entsprechend der obigen Ausführungen unter Beibehaltung ihrer Anfangsbuchstaben ersetzt. Letzteres ist notwendig, damit die Akronymeigenschaft erhalten bleibt. Es werden nach Möglichkeit phonetisch ähnliche Begriffe wie im ursprünglichen Akronym verwendet, um eine zusätzliche Diskrepanz zwischen

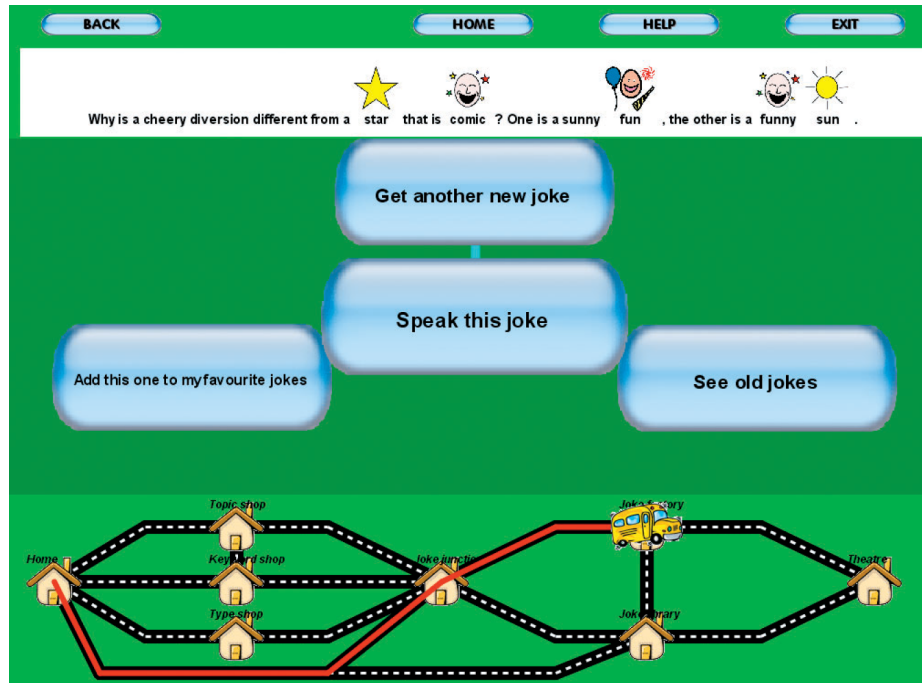


Abbildung 3. Kindgerechte Benutzerschnittstelle von STANDUP. (Abbildung aus [7])

der Ähnlichkeit zu Bekanntem und den neuen, widersprüchlichen Ausdrücken zu erzeugen. Als Rückfalllösung kann auf ein manuell zusammengestelltes Wörterbuch mit kontroversen oder überspitzenden Worten zurückgegriffen werden. Ein Beispiel für ein modifiziertes Akronym findet sich in Abbildung 4.

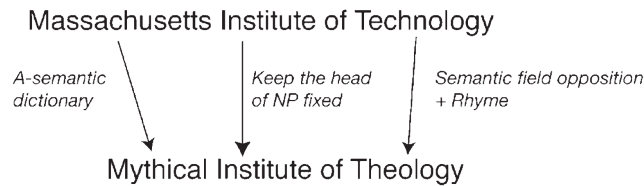
Die Generierung neuer Akronyme verläuft analog zu der beschriebenen Modifikation. Anstelle des Parsens wird hier eine Nominalphrase als WordNet-Eintrag vorgegeben und die Parsergrammatik zur Erstellung eines sprachlich korrekten Gerüsts benutzt.

Die Qualität der Akronyme wurde in Befragungen mit jeweils 20 Studenten ermittelt. Dabei wurden modifizierte Akronyme in mehr als 70 % der Fälle als *lustig genug* (im Durchschnitt  $>2.75$  auf einer Skala von 1 [*nicht lustig*] bis 5 [*sehr lustig*]) bewertet. Keine der abgewandelten Abkürzungen wurde im Mittel als *nicht lustig* eingestuft. Bei neu generierten Akronymen wurden 53 % als *lustig genug* empfunden, in einer Kontrollgruppe mit grammatikalisch passenden, aber ansonsten zufällig generierten Worten, waren es 8 %.

### 3.4 Modellbasierte Erkennung von Humor

Während sich die bisher vorgestellten Systeme mit der Erzeugung von Humor befasst haben, wird im Folgenden ein Ansatz beschrieben, wie humorvolle Äußerungen durch Rechnersysteme *erkannt* werden können [19]. Gegenstand dieser





**Abbildung 4.** Beispiel: Modifikation eines Akronyms. Während der Kopf der Nominalphrase (*Institute*) beibehalten wird, um kein gänzlich mit dem ursprünglichen unzusammenhängendes Akronym zu erhalten, wird bei den übrigen Bestandteilen auf die Rückfalllösung (*A-semantic dictionary*) bzw. auf WordNet als Quelle zurückgegriffen. (Abbildung aus [16])

Vorgehensweise sind humorvolle Einzeiler, die beispielsweise folgende Gestalt haben:

*Beispiel 3. Beauty is in the eye of the beer holder.*

Das Ziel ist eine Unterscheidung von ähnlich aufgebauten, aber nicht humorvollen Aussagen, wie dem ursprünglichen Sprichwort:

*Beispiel 4. Beauty is in the eye of the beholder.*

Um eine solche binäre Klassifikation (humorvoll vs. nicht humorvoll) zu erreichen, werden zwei Ansätze getestet und kombiniert. Zum einen werden als humortypisch vermutete stilistische Merkmale (Antinomie, Alliteration, *Slang*) in einem Entscheidungsbaumverfahren genutzt, zum anderen eine Textklassifikation mittels *Naive-Bayes*-Klassifikator und einer *Support Vector Machine (SVM)*. Die Datenbasis bilden 16 000 humorvolle Einzeiler und jeweils 4 000 nicht witzige Sätze von Überschriften aus Artikeln der *Reuters*-Nachrichtenagentur, Sprichworten, dem *British National Corpus (BNC)* und dem Projekt *Open Mind Common Sense (OMCS)*, einer Sammlung von Sätzen über Weltwissen. Beispiele neben dem oben genannten Sprichwort sind:

*Beispiel 5. Reuters: Oil prices slip as refiners shop for bargains.*

*Beispiel 6. BNC: The train arrives three minutes early.*

*Beispiel 7. OMCS: A file is used for keeping documents.*

Die Schwellenwerte, die festlegen, ab welcher Ausprägung ein Stilmittel auf das Vorliegen von Humor schließen lässt, werden mit einer Auswahl von 1 000 als humorvoll bzw. nicht humorvoll charakterisierten Sätzen als Trainingsdaten bestimmt. Die Ergebnisse einer anschließenden Evaluation mit den übrigen Daten sind in Tabelle 5a zusammengefasst. Auffällig ist, dass Alliterationen für sämtliche Datensätze der beste Indikator bezüglich Humor zu sein scheint. Insgesamt sind die betrachteten stilistischen Unterschiede zwischen den Textarten mit Ausnahme

					One-liners				
						Reuters	BNC	Proverbs	OMCS
					Naive Bayes	96.67%	73.22%	84.81%	82.39%
					SVM	96.09%	77.51%	84.48%	81.86%
(b) Erkennung mit Textklassifikation									
					One-liners				
						Reuters	BNC	Proverbs	OMCS
					Combination	96.95%	79.15%	84.82%	82.37%

One-Liners				
Heuristic	Reuters	BNC	Proverbs	OMCS
Alliteration	74.31%	59.34%	53.30%	55.57%
Antonymy	55.65%	51.40%	50.51%	51.84%
Adult slang	52.74%	52.39%	50.74%	51.34%
All	76.73%	60.63%	53.71%	56.16%

(a) Erkennung mit Stilmerkmalen

(c) Erkennung mit Kombination

**Abbildung 5.** Erkennungsraten bei Verwendung unterschiedlicher Methoden. (Abbildungen aus [19])

der Reuters-Überschriften jedoch gering. Unter Betrachtung der Vorüberlegungen in Abschnitt 2 erscheint dies nicht weiter verwunderlich: Während Inkongruenz durch die Wahl von Stilmitteln verstärkt werden *kann*, ist es offensichtlich nicht *hinreichend*, Gegensätze oder Obszönitäten einzuflechten, um Humor zu erzeugen.

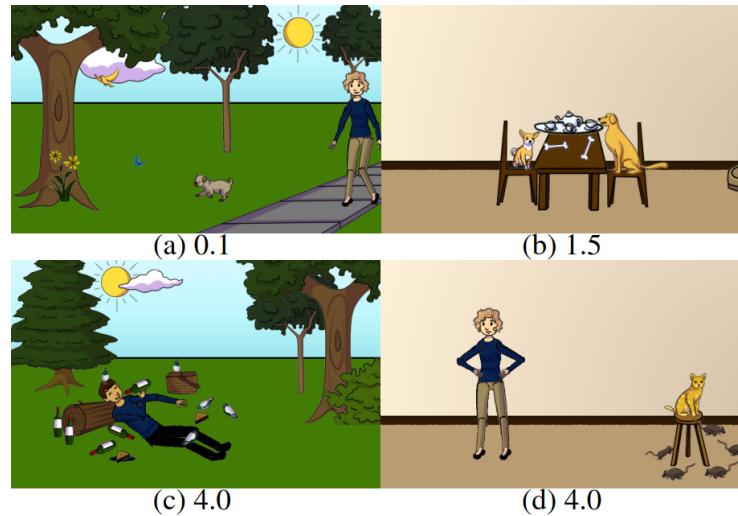
Die inhaltsbasierte Textklassifikation hingegen zeigt, wie in Tabelle 5b erkennbar, eine höhere Erkennungsrate. Hier sind ebenfalls die Reuters-Überschriften am besten von witzigen Einzeilern zu unterscheiden. Auch die Sprichworte, die sich stilistisch kaum von den Einzeilern unterscheiden, können so zuverlässiger klassifiziert werden. Deutlich wird dies an den eingangs genannten Beispielen: Die Levenshtein-Distanz zwischen dem Witz und dem Sprichwort beträgt lediglich 3, während semantisch ein großer Unterschied zwischen *beer holder* und *beholder* auszumachen ist. Eine dementsprechend geringe Verbesserung zeigt auch eine Kombination beider Vorgehensweisen, wie in Tabelle 5c dargestellt ist; einzig Daten aus dem BNC können so mit geringfügig höherer Erkennungsrate klassifiziert werden.

### 3.5 Visual Humor

Die bisher beschriebenen Ansätze haben sich ausschließlich mit Humor in textueller Form befasst. In diesem Abschnitt wird ein Verfahren vorgestellt, mit dem humorvolle Bilder erkannt und abgewandelt werden können, sodass sie mehr oder weniger lustig als ursprünglich sind [20]. Gegenstand der Betrachtung sind *Clipart*-Szenen, in denen vor unterschiedlichen Hintergründen etwa 150 verschiedene Objekte frei angeordnet werden können, darunter verformbare menschliche Figuren, Tiere und Alltagsgegenstände.

Als Datenbasis für die Einschätzung der Lustigkeit dienen 6 400 von *Crowdworkern*<sup>2</sup> erstellte Szenen, die von wiederum anderen Arbeitern auf einer Skala von 1 bis 5 entsprechend ihres Humorgehaltes bewertet wurden. Beispiele sind in Abbildung 6 zu sehen. Der Witzigkeitswert (*funniness*) einer Szene entspricht dabei dem Durchschnitt der Bewertungen durch 10 Crowdworker. Ein zweiter Datensatz bestehend aus rund 15 000 weiteren Szenen wird für die Abwandlung

<sup>2</sup>Selbstständige Arbeiter auf speziellen Online-Plattformen.



**Abbildung 6.** Von Testpersonen bestimmter Witzigkeitswert verschiedener Clipart-Szenen. (Abbildung modifiziert nach [20])

der Witzigkeit genutzt. Dazu wurden die Crowdworker beauftragt, humorvolle Szenen aus dem ersten Datensatz *weniger* lustig zu machen, indem sie Objekte gegen andere austauschen. Dabei wurden sie angehalten, so wenige Veränderungen wie möglich zu tätigen und den Grundaufbau der Szene beizubehalten. Auf diese Weise sollen Erkenntnisse gewonnen werden, welche spezifischen Objekte und Kategorien von Objekten zu humorvollem Empfinden führen und unter welchen Umständen dies der Fall ist.

Um ein Modell für die *Einschätzung* der Witzigkeit einer Szene zu trainieren, werden verschiedene Eigenschaften genutzt: die Anzahl Instanzen von Objekten unterschiedlicher Objektkategorien, die Position und die Gesamtzahl Objekte in der Szene. Mittels eines *Support-Vector-Regressors* wird das Modell unter Bezug auf die Bewertungen der Crowdworker angepasst, um den Witzigkeitswert von Szenen durch die oben genannten Merkmale zu prognostizieren.

Für die *Veränderung* der Witzigkeit wird ein mehrschichtiges Perzeptron trainiert, das für jedes Objekt einer Szene entscheidet, ob es weiterhin erhalten bleibt oder durch ein anderes Objekt ersetzt wird, das zu einem gesteigerten oder verminderten Humorgehalt führen soll. Dazu werden Eigenschaften auf der Objektenebene genutzt: Zum einen der Kontext, in dem das Objekt *üblicherweise* auftritt, also welches Vorhandensein andere Gegenstände das Auftreten wahrscheinlicher machen, zum anderen die Entfernung zu den übrigen Objekten der Szene; auf diese Weise kann eine Art (In-)Kongruenz repräsentiert werden. Als Trainingsdaten für das Modell dienen die Informationen über die von Crowdworkern ausgetauschten Objekte der jeweils betrachteten Szene. Um passende Ersatzobjekte zu bestimmen, die eine Szene mehr oder weniger witzig machen, wird jeweils ein

Modell auf Grundlage der witzigen Szenen, respektive der von Crowdworkern weniger witzig gemachten Entsprechungen des zweiten Datensatzes, angepasst.

Bei der Evaluation des Modells für die Vorhersage der Witzigkeit wird ein durchschnittlicher relativer Fehler von 0.24 erreicht. Im Vergleich wird der durchschnittliche Witzigkeitswert herangezogen, den die Crowdworker für die Testdaten vergeben haben. In diesem Fall beträgt der relative Fehler im Mittel 0.32. Die Modelle zur Veränderung der Witzigkeit werden getestet, indem jeweils 10 Testpersonen aufgefordert werden, die ursprünglichen und die abgewandelten Szenen hinsichtlich ihrer Lustigkeit nach der bereits beschriebenen Skala zu bewerten. Im Falle einer gewünschten Verringerung der Witzigkeit liegt die Bewertung der Ausgangsszenen durchschnittlich bei 2.69, nach Bearbeitung sinkt diese auf 1.64. Im umgekehrten Szenario, der Steigerung der Witzigkeit, kann eine durchschnittliche Wertung von 2.14 erreicht werden. Bei einem direkten Vergleich zwischen den ursprünglichen, von Menschen erstellten lustigen Szenen und denen, die durch das Modell generiert wurden, empfinden die Versuchspersonen letzteres in 28 % der Fälle als lustiger. Beispiele für die Abwandlung der Witzigkeit finden sich in den Abbildungen 7 und 8.

### 3.6 Weitere Arbeiten

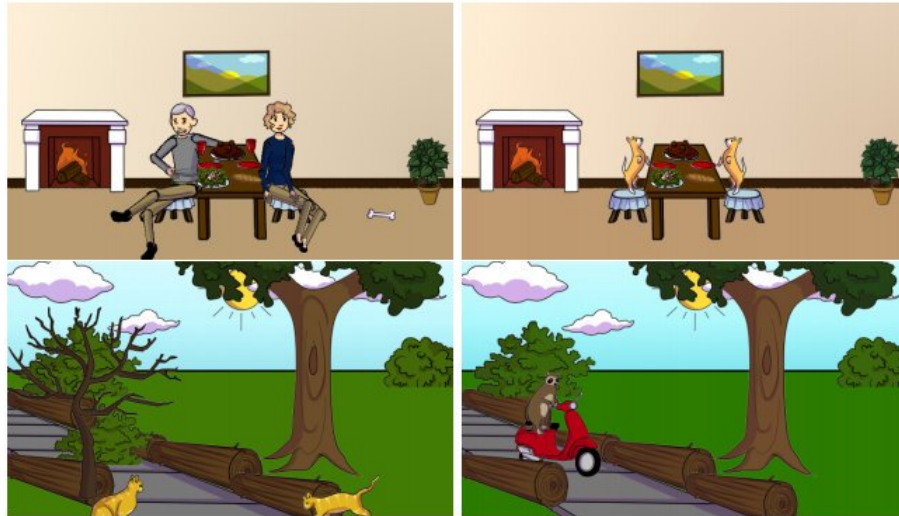
Auch wenn mit den vorausgehenden Abschnitten der Versuch unternommen wird, einen möglichst vielfältigen Überblick über das Feld des *Computational Humors* zu geben, ist dies durch den begrenzten Umfang der Ausarbeitung nur eingeschränkt möglich. Weitere Arbeiten, die sich mit anderen Aspekten von Humor befassen, sollen im Folgenden kurz angerissen werden.

Soziale Medien bieten eine große Datenbasis, um menschliche Kommunikationsweisen zu untersuchen. Besonders prägnant präsentieren sich diese auf dem Kurznachrichtendienst *Twitter*. So beschäftigen sich Untersuchungen mit der automatischen Erkennung von Ironie [21] und Sarkasmus [22] mit Methoden des maschinellen Lernens in den öffentlich zugänglichen Nachrichten. Eine Besonderheit der letzteren Arbeit ist eine Anreicherung mit Daten zu den *Beziehungen* der Kommunikationspartner zueinander, statt ausschließlich linguistische Merkmale zu verwenden.

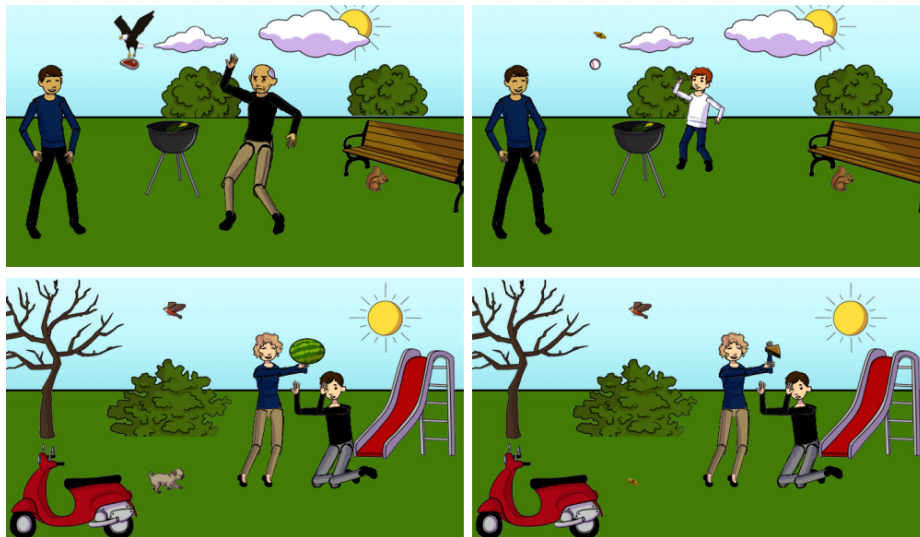
Eine weitere Arbeit setzt direkt bei am *Incongruity-Resolution*-Modell an und versucht, verschiedene Dimensionen von Inkongruenz zu formalisieren und evaluieren: Mehrdeutigkeit und unterschiedliche Blickwinkel auf gegebene Situationen. Darüber hinaus wird untersucht, wie mit dem gewonnenen Modell witzige Elemente in Wortspielen identifiziert werden können. [23]

## 4 Zusammenfassung und Ausblick

In dieser Ausarbeitung wurde ein Überblick über das vergleichsweise junge Feld des *Computational Humors* gegeben. Dabei wurden zunächst interdisziplinäre Perspektiven zu den theoretischen Grundgerüsten von Humor umrissen. Es wurden Schwierigkeiten verdeutlicht, eine eindeutige und umfassende Definition des



**Abbildung 7.** Wandlung von unlustigen Bildern (links) zu witzigen Bildern (rechts). Während in den Ursprungsszenen alltägliche Situationen dargestellt sind, wird durch Inkongruenz (Hunde am Esstisch; Waschbär auf Motorroller) Humor erzeugt. (Abbildung aus [20])



**Abbildung 8.** Wandlung von witzigen Bildern (links) zu unlustigen Bildern (rechts). Der Adler mit gestohlenem Steak wird durch einen Ball und einen Schmetterling ersetzt. Weiterhin wird Kongruenz erzeugt, indem ein Junge den älteren Mann substituiert und so eine Erklärung für die Anwesenheit des Balles bietet. In der unteren Szene wird u. a. die Melone durch eine Axt ersetzt, was kongruent zu dem verängstigten Mann ist (und der lachenden Dame eine gewisse Psychopathie attestiert). (Abbildung aus [20])

Begriffs zu geben, da unter der Bezeichnung Humor oftmals eine Sammlung verschiedener menschlicher Ausdrucksformen verstanden wird. Darüber hinaus wird deren Zurechnung durch individuelle und kulturelle Unterschiede beeinflusst. Mit dem *Incongruity-Resolution*-Modell wurde eine der vorherrschenden Theorien im Bereich der Humorforschung beschrieben, die ein weitgehend vereinheitlichendes, wenn auch nicht allumfassendes Rahmenwerk dessen bietet, was weithin als humorvoll angesehen wird.

Auch wenn das *Incongruity-Resolution*-Modell von individuellen und gesellschaftlichen Differenzen abstrahiert, wurde dargestellt, dass eine Übertragung auf die maschinelle Ebene mit Schwierigkeiten verbunden ist. Umfangreiches semantisches Wissen und Schlussfolgern ist notwendig, um Inkongruenz zu erkennen, einzuordnen und aufzulösen. Als ein wichtiges Werkzeug im Umgang mit diesem Problem wurde das lexikalisch-semantische Netz WordNet vorgestellt. Dieses bildet eine zentrale Komponente in den vorgestellten Generatoren für lustige Rätsel, JAPE und STANDUP. Während auf diese Weise qualitativ annehmbare Wortspiele hervorgebracht werden können, wurde die starke Abhängigkeit von menschlicher Zuarbeit in Form von manuell erstellten Witzschemata angemerkt. Als weiteres Beispiel für ein um WordNet und dessen Erweiterungen entwickeltes System wurde der Akronym-Generator HAHAAcronym beschrieben, der auch ohne weitere menschliche Intervention humorvolle Akronyme aus bereits vorhandenen oder von Grund auf neu erstellen kann.

Weiterhin wurde ein Ansatz beschrieben, mit dem Humor, hier in der Form von lustigen Einzeilern, von Rechnersystemen durch Techniken des maschinellen Lernens *erkannt* und von nicht witzigen Texten abgegrenzt werden kann. Dabei wurde herausgestellt, dass die Ergebnisse dieses Verfahrens im Einklang mit dem *Incongruity-Resolution*-Modell stehen: Humor *kann* durch rhetorische Stilmittel auf der syntaktischen, semantischen oder phonetischen Ebene verstärkt werden, ist aber nur zu einem Teil durch diese erklär- und erkennbar; erst durch die Betrachtung der *inhaltlichen* Aspekte kann zwischen *lustig* und *nicht lustig* zuverlässiger differenziert werden.

Als Beispiel eines Verfahrens, das sich *nicht* mit sprachbasiertem Humor beschäftigt, wurde die modellbasierte Einschätzung und Veränderung der Witzigkeit von *Clipart*-Szenen dargestellt. Während sich Verbesserungen gegenüber naiven Vorgehensweisen zeigten, konnte ein menschliches Niveau bei der Erzeugung von Humor im Allgemeinen nicht erreicht werden.

Auch wenn sich diese Aussage resümierend auf den gesamten betrachteten Bereich des *Computational Humors* verallgemeinern lässt, konnten wichtige Grundlagen erarbeitet werden. Diese bekommen umso größere Bedeutung, wenn das lückenhafte Verständnis von *Humor an sich* zugrunde gelegt wird. Denkbar ist, dass mit wachsenden Fortschritten in Bereichen wie dem *Natural Language Processing* und dem maschinellen Sehen weitaus komplexere Humorformen als kindgerechte Wortspiele erfasst werden können – und dies in einer wechselseitigen Beziehung auch dem *Menschen* ein besseres Verständnis seines eigenen Humors ermöglicht.

Eine Maschine mit menschenähnlichem Sinn für Humor bleibt mutmaßlich zunächst eine ferne Zukunftsvision. Als essenzieller Bestandteil menschlichen Erlebens und Sozialisierens ist dieses bisher nur in Ansätzen ergründete Themenfeld dennoch von Wichtigkeit; sollen intelligente künstliche Systeme einmal menschenähnliches Bewusstsein erlangen oder zumindest glaubhaft diesen Anschein erwecken können, sollte diese Fähigkeit nicht ausgespart werden [24]:

BARTENDER GUINAN: *Being able to make people laugh, or being able to laugh, is not the end all and be all of being human.*

ANDROID DATA: *No, but there is nothing more uniquely human.*

## Literatur

1. Mulder, M.P., Nijholt, A.: Humour research: State of the art. Technical Report 34, CTIT Technical Report Series (2002)
2. Ruch, W., ed.: The Sense of Humor: Explorations of a Personality Characteristic. Walter de Gruyter (1998)
3. Goel, V., Dolan, R.J.: The functional anatomy of humor: Segregating cognitive and affective components. *Nature Neuroscience* **4**(3) (2001) 237–238
4. Wild, B., Rodden, F.A., Grodd, W., Ruch, W.: Neural correlates of laughter and humour. *Brain* **126**(10) (2003) 2121–2138
5. Hauser, S.: Wie Kinder Witze erzählen: Eine linguistische Studie zum Erwerb narrativer Fähigkeiten. Peter Lang (2005)
6. Shahaf, D., Amir, E.: Towards a theory of ai completeness. In: AAAI Spring Symposium: Logical Formalizations of Commonsense Reasoning. (2007) 150–155
7. Binsted, K., Nijholt, A., Stock, O., Strapparava, C., Ritchie, G., Manurung, R., Pain, H., Waller, A., O’Mara, D.: Computational humor. *IEEE Intelligent Systems* **21**(2) (2006) 59–69
8. Gudehus, C., Eichenberg, A., Welzer, H.: Gedächtnis und Erinnerung: Ein interdisziplinäres Handbuch. J.B. Metzler (2016)
9. Miller, G.A.: Wordnet: A lexical database for english. *Communications of the ACM* **38**(11) (1995) 39–41
10. Hamp, B., Feldweg, H.: Germanet – a lexical-semantic net for german. In: Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications. (1997)
11. Naber, D.: Openthesaurus: ein offenes deutsches wortnetz. *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen: Beiträge zur GLDV-Tagung in Bonn* (2005) 422–433
12. Binsted, K.: Machine Humour: An Implemented Model of Puns. PhD thesis, University of Edinburgh (1996)
13. Manurung, R.: Interactive computer generation of jokes for language skill development. Lecture Slides (2003)
14. Binsted, K., Pain, H., Ritchie, G.D.: Children’s evaluation of computer-generated punning riddles. *Pragmatics & Cognition* **5**(2) (1997) 305–354
15. Manurung, R., Ritchie, G., Pain, H., Waller, A., O’Mara, D., Black, R.: The construction of a pun generator for language skills development. *Applied Artificial Intelligence* **22**(9) (2008) 841–869
16. Stock, O., Strapparava, C.: Hahacronym: Humorous agents for humorous acronyms. *Humor* **16**(3) (2003) 297–314

17. Magnini, B., Cavaglia, G.: Integrating subject field codes into wordnet. In: Proceedings of the Second International Conference on Language Resources and Evaluation. (2000) 1413–1418
18. Strapparava, C., Valitutti, A.: Wordnet-affect - an affective extension of wordnet. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation. (2004) 1083–1086
19. Mihalcea, R., Strapparava, C.: Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence* **22**(2) (2006) 126–142
20. Chandrasekaran, A., Vijayakumar, A.K., Antol, S., Bansal, M., Batra, D., Zitnick, L.C., Parikh, D.: We are humor beings: Understanding and predicting visual humor. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2016) 4603–4612
21. Barbieri, F., Saggion, H.: Modelling irony in twitter. In: European Chapter of the Association for Computational Linguistics. (2014) 56–64
22. Bamman, D., Smith, N.A.: Contextualized sarcasm detection on twitter. In: The International AAAI Conference on Web and Social Media. (2015) 574–577
23. Kao, J.T., Levy, R., Goodman, N.D.: The funny thing about incongruity: A computational model of humor in puns. In: Proceedings of the Annual Meeting of the Cognitive Science Society. (2013)
24. Paramount Pictures: Star trek: The next generation – the outrageous okona (1988)