

日本語におけるネガティブ言語変換システムの開発

須永直人^{†1} 星野貴仁^{†2} 舘岡大貴^{†3}

本研究は、日本語におけるコミュニケーション、主にネガティブ文字情報（ネガティブ言語）によって対人関係が悪化するリスクを低減させ、コミュニケーションを楽しくする技術開発を目的とする。SNS等の日本語コミュニケーションにおいても文字情報の役割が増加する中で、本来非言語コミュニケーションや文脈から伝わるはずの情報量が制約されることで生じる対人関係のリスクに着目し、ポジティブに自動変換するシステムを提案する。

Proposing a system for converting negative Japanese language into positive Japanese language

NAOTO SUNAGA^{†1} TAKAHITO HOSHINO^{†2}
DAIKI TATEOKA^{†3}

Depending chiefly on the negativity of textual information (negative language), personal relations can worsen. This paper aims to develop techniques that reduce this risk and consequently make communication more enjoyable. As the number of forms of Japanese language textual information on SNS (Social Networking Services) and other similar services increases, the risk of miscommunication and damage to personal relationships that naturally arises from the limited amount of contextual information communicated non-verbally through such platforms also increases. Focusing on this problem a system to automatically convert negative language into positive language is proposed.

1. はじめに

近年、ITの進歩に伴いSNSツールが急速に社会に浸透し、普及している。たとえば、代表的なSNSツールであるLINE[1]やTwitter[2]等は、言語記号による意味を手がかりとする言語(verbal)コミュニケーション情報のやり取りが大部分を占めるツールであり、身体動作などに代表される非言語(non-verbal)コミュニケーション情報は他者との対面状況(face-to-face)に比べ重要な情報とはならない。こうした内容はこれまでのコミュニケーション研究からも示唆されており[3]、SNSツールでは、主にテキストに代表される言語コミュニケーションの重要性を指摘することができる。対面状況でのコミュニケーションと同様に、SNSツール上においても、受け手が送り手から受け取る言語情報はポジティブなものばかりではなく、「辛い」や「悲しい」といったネガティブな言語情報も含んでいることが想定される。たとえば、ネガティブな言語情報を受け取った場合、コミュニケーションを継続したいという動機が消失してしまう可能性もある。そこで本稿では、こういったSNSツール上でのネガティブ言語情報を「ネガティブ言語」と表し、Entertainment Computing技術を用いて、より楽しく、コミュニケーションを続けたいと思えるような「日本語におけるネガティブ言語変換システム」を開発・提案することを目的とする。

1.1 認知的不協和とネガティブ言語

フェスティンガー(Festinger, L.) [4]は、認知に不協和が発生した場合、人にとってその不協和は不快であるため、不協和を低減したり回避したりしようとする試みを実行することを明らかにし、「認知的不協和」と定義した。SNSツール上においても認知的不協和は発生する可能性があり、認知的不協和が発生したと仮定すると、自己にとって不都合な認知を変化させようとする働きがSNSによるコミュニケーションにおいても起こりうると思われる。たとえば、SNSツール上で送り手からネガティブ言語を含むメッセージを受け取ると、受け手に送り手との持続的で良好な関係を望む動機づけがある限り、送り手との関係性に影響が現れないよう注意を払い、メッセージを返信すると考えられるだろう。一方で、Adamら[5]が指摘するように、他者と同様の感情を自己が知覚してしまう現象である「情動伝染(emotional contagion)」がウェブコンテンツの閲覧のような非直接的コミュニケーションにおいても起こることを考慮すれば、受け手が送り手のネガティブ言語メッセージを受け取った場合、受け手もネガティブな感情を認知してしまうと考えられる。にもかかわらず、送り手との良好な関係性の持続に妥当なポジティブ言語を含んだメッセージを返信する行為は、自身の実際のネガティブな感情と矛盾しているため、認知的不協和を引き起こすと考えられる。この知覚したネガティブな認知を無理にポジティブ言語で

†1 (株式会社須永総合研究所)

†2 (株式会社 SHP)

†3 (株式会社 SHP)

表出する行動は、対人関係におけるリスク回避的行動であると言えるが、同時に認知的不協和は解消されないことからストレスとなり得る行動であるとも言える。

本稿では、このような状況に直面した受け手を対象とし、対人関係リスクの回避と受け手のストレス低減を両立しながら、SNS等の言語コミュニケーションをより「楽しく」することのできるネガティブ言語変換システムを提案する。ネガティブ言語変換システムは、Entertainment Computing技術として、ネガティブ言語メッセージによって送り手にネガティブな認知を抱いている受け手が、自身のネガティブな感情を否定したり、偽ったりすることなく、ポジティブ言語表現に変換する行為を自動化することのできる、コミュニケーション支援ツールである。受け手がこのシステムを利用することで、自身の感情と関係なく、メッセージを自動でポジティブな内容に変換してくれるため、認知的不協和の発生頻度の低減が期待できる。認知的不協和の発生を低減できれば、受け手がコミュニケーションを続ける動機の低減を抑え、「より楽しく」コミュニケーションを続けることができるだけでなく、対人関係の悪化や、自身の認知的不協和による不快感などのストレスを低減することも期待される。

1.2 言語変換とユーモア

これまでネガティブ言語と精神的な影響に着目した研究は、多数行われてきた。大家ら[6]は、SNSツール上でのネガティブ感情表現の影響を受けてしまう読み手を対象とし、精神的な影響の緩和を目的に、テキストをネガティブフレーズからポジティブフレーズへと換言するリフレーミングの技法を用いたフィルタリング手法の提案と、それを実現するインタフェースを実装した。このシステムを利用することで読み手は、気分に応じて換言レベルを変更し、フレーズを変更できる。また、リフレーミングの技法を応用し、多数のネガティブ言語に対応するポジティブ言語を掲載しているスマートフォンアプリ「ネガボ辞典」[7]は、ネガティブ言語をポジティブ言語に変換することによって印象の変化をサポートするシステムと言える。たとえば、「ばか」というネガティブ言語に対応するポジティブ言語は「素直」や「天才」と表記して印象の変化を可能にしている。リフレーミングは、個人の認知的枠組みを変えて態度対象への印象を変化させる技法であるため、ネガティブ言語からポジティブ言語への換言の際の指針として有効であると推測できる。しかし、リフレーミングの技法のみを応用して言語変換を行うことには、以下のような問題点が指摘されよう。すなわち、その理由のひとつは、送り手と受け手の相互作用による言語コミュニケーションを対象にした場合、送り手と受け手の関係性の変化が生じる可能性が出てくる。たとえば、上述した研究やシステムでは、受け手のネガティブ感情によるストレスの低減を主な目的と

しているため、こうした対人関係の変化に関しては十分に配慮されたシステムとは言えない。本稿ではこうした対人関係の変化やそれに伴う対人関係リスクに着目し、対人関係が悪化するリスクを回避するという、本来であれば後ろ向きでネガティブな目的を、Entertainment Computing技術を活用することで、コミュニケーションや対人関係を「楽しく」自ら継続したいと思えるような積極的でポジティブなコミュニケーションを支援するツールによって、昨年のEC2016招待講演においてAnton Nijholt氏が言及されていた「赤信号を楽しく待つ技術」のような構造を「不快なコミュニケーションを楽しむ技術」として提案できるのではないかと考えている。

とは言え、受け手にとってのポジティブ言語が、送り手にとっても全てポジティブに受け取られるとは限らない。たとえば、「あなたはばかだ」というメッセージを「あなたは天才だ」に変換し送り手に返信した場合、送り手がネガティブな認知を持っていれば「皮肉」と理解され、その対人関係が悪化してしまうリスクが懸念される。このようなコミュニケーションパターンの課題は、リフレーミングの応用のみでは解決が難しいと言える。そこで、本研究においてはポジティブ言語に変換されたメッセージを受け取る側（本来のメッセージの送り手）がそのメッセージによってポジティブな感情を知覚することで低減されると考える。

RomeroとCruthirds[8]は「個人、グループ、組織におけるポジティブな感情や認知を生み出す楽しいコミュニケーション」として、ユーモアを広義に取り扱っている。また、Holmes[9]は、職場におけるユーモアの効果に着目し、ユーモアがストレス対処やチームワークの促進に効果的であると論じている。Vinton[10]はユーモアがメンバー間の地位の差異により生じる衝突を抑え、緊張の緩和に影響することを指摘している。一方、上野[11]はユーモアを「おかしさやおもしろさという心的現象を示すもの」と定義し、ユーモアとストレス緩和の実証研究結果に一貫性がないことから、ユーモア現象全体を3種類に分類する必要性を指摘している。これらの研究から、ユーモアを持つ働きはポジティブな感情を生起させる手段として有用であることが示唆されるものの、ユーモア現象を一律に扱うことの困難さも同時に示唆されている。こうした結果から、本研究においては対人関係の悪化リスクを低減する目的でユーモアを用いる際、上野[11]のユーモア分類を参考にすることでユーモア現象をより取り扱いやすいよう工夫を行った。使用した分類である3種類のユーモアとは、自己や他者を楽しませるために表出する「遊戯的ユーモア」、自己や他者を攻撃するために表出する「攻撃的ユーモア」、自己や他者を励まし、許し、心を落ち着けるために表出する「支援的ユーモア」である。上野[11]は、この3種類のユーモアのうち、精神的健康状態と関連するユーモアは支援的ユーモアであると考察している。宮戸ら[12]は支援的ユーモアの研究に

において、「ユーモアの生起には、楽しさを求める享乐的な姿勢が求められるが、そこに自己を客観視し、洞察しようとする傾向が加わって初めてユーモアの支援的効果がもたらされると推定される。単に楽しさを求める享乐的な姿勢だけでは、ユーモアの支援的効果はもたらされない。支援的ユーモアは自己や他者を精神的に励まし、支えようとする意図性が高いと考えられる。」とし、支援的ユーモアの独立性を指摘した。塚脇ら[13]は、支援的ユーモアが他のユーモアとは異なり、表出する形態と動機を同一で扱い分類している問題を指摘し、支援的ユーモアを「自虐的ユーモア」と捉えた。この問題に関しては、上野[14]も刺激として支援的ユーモアと遊戯的ユーモアの分別は難しいと述べており、「自虐的ユーモア」は支援的ユーモアの刺激としての側面においては改善の役割を果たしたと言えるだろう。しかし、ユーモアを表出する動機によってはダジャレなどの言葉遊びも支援的ユーモアになり得るという点はより注意深く考察が必要のように思える。たとえば、遊戯的ユーモアにおいても、受け手は送り手に対して「親しみやすさ」を感じる研究(牧野[15])や、信号無視をしてもピエロの格好で笑わせながら通るとクラクションを鳴らされない実験(Dworkin & Efran[16])等の研究からも示唆されるように、支援的ユーモアの機能の一部とも考えられる、怒り等のネガティブ感情を抑制する効果があると考えられる。

以上の先行研究を踏まえ、本研究ではユーモアの中でも特に、支援的ユーモアと遊戯的ユーモアに対人関係が悪化するリスクを低減する効果があると考え、ネガティブ言語変換システムの指針として両ユーモアの要素を取り入れた。

2. ネガティブ言語変換システムの構成

2.1 変換システムの使用状況

本項目では、1章で提案した変換システム手法の詳細を、図を用いて述べていく(図1)。

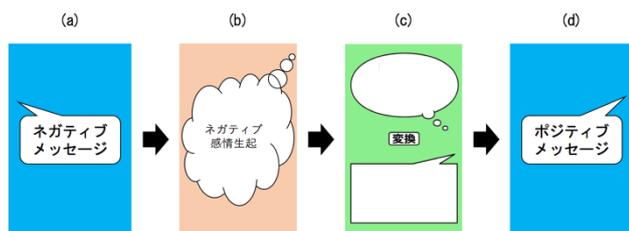


図1 変換システム使用の流れ

本システムは、メッセージの送り手と受け手による対人場面を想定し、ネガティブ感情を知覚した受け手を対象としたシステムである。まず、SNS ツール上で受け手は送り手からネガティブ言語を受信する(a)。受け手は送り手のネガティブ言語に影響されネガティブな感情を認知する(b)。ここで、自然に関係を維持するためのメッセージを返信しようとする場合、認知的不協和が発生するリスクが生じる。一方、本システムを利用すれば、受け手は心的な

負荷を感じずに、ネガティブ言語をポジティブ言語へと変換することができる(c)。変換されたポジティブ言語を参考にし、送り手へポジティブメッセージの返信を行う(d)。受け手に取って、自らのネガティブ情報をポジティブに変換したのはあくまで本システムであり、自分自身ではない。自身は本システムが変換した内容を相手に伝えているだけであり、仮にシステムによるアウトプットが自己の感情と矛盾するメッセージであったとしても、直接的な自己矛盾は生じないことから認知的不協和が発生するリスクを低減することが期待される。言い換えれば、本システムによって、ネガティブ言語の刺激が外在化されることによって、心的負荷を軽減する支援が期待される。本稿では、この(c)に該当するシステムの設計及び機械学習による人工知能を用いたコンピュータプログラムの開発並びにユーザーインターフェースの開発を行った。

2.2 変換システムの構成

本システムの中核である変換プログラムにおいては、各ネガティブ言語とそれぞれに対応するポジティブな表現とのセットを教師データとして学習し、ネガティブな文章を読み込むことで、適切なポジティブな表現に置き換えて表示するプログラムを構築する。本システムの重要な点は、単に辞書から抜き出して変換するのではなく、システムに機械学習による人工知能を取り入れることで類似するネガティブな表現をシステム自ら判断し、ポジティブな表現に変換を行うことが可能になる点である(図2)。

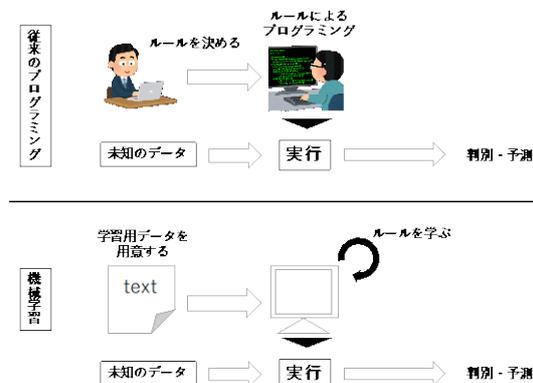


図2 従来のプログラミングと機械学習によるAI

判別の手法に機械学習による人工知能を用いることで、未知のデータに対する変換精度の向上が期待できる。とは言え、コミュニケーションは多様であり、かつまた、認知における個人差の影響も大きいと考えられる。そこで、本システムの開発にあたって、ネガティブ言語の定義を明確化した上で、まずは場面特異的な以下の2状況に絞り込み、ポジティブ変換のプログラムを構築した。本研究で想定した場面状況は、関係継続に対する動機が比較的高いと考えられる友人同士のコミュニケーションと関係の持続が一般

に大きなライフイベントに直結していると考えられる恋人同士の恋愛関係コミュニケーションに限定し、辞書の作成および変換システムの構築を行った。

3. 変換のメカニズム

インターネット上に存在している主に SNS におけるネガティブ言語候補と言える日本語表現を抽出し、各ネガティブ言語に対応するポジティブ表現のリスト化を行った。その後、元になったネガティブ表現を形態素に分解し、対応するポジティブ表現を入力した教師データを作成した。教師データを用いて機械学習を行い、ネガティブな表現とそれに対応するポジティブな表現に結びつく特徴量を抽出し、自動で変換を行うプログラムを構築した。開発については、データのクローリングから機械学習、表示のインタフェースまで Python 言語を用いて開発を行った。

4. 開発手順

まずはPythonを用いてTwitterに接続し、一般に公開されている投稿の中からネガティブなツイートのデータを収集した。その際、本研究で取り上げる場面特異的な状況と親和性が高いと考えられる単語、たとえば「バカ」、「ウザい」といったネガティブ単語を、キーワードとシクロリングを行なった (図3)。

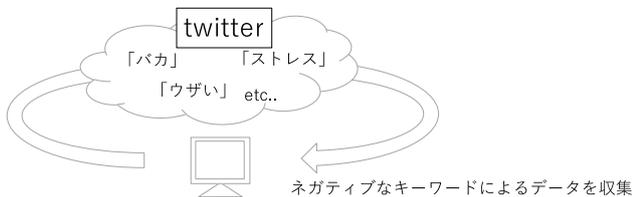


図3 データ収集イメージ

次に、クローリングしたネガティブ表現から本研究の場面特異的な状況に合致する表現を残した上で形態素に分解し、対応するポジティブな表現を正解データとして入力した教師データを作成し、教師データを用いて機械学習を行った。(図4)

元データ	学習用データ	正解ラベル
お前は口が悪い	お前, は, 口, が, 悪い	あなたはとても素直ですね
⋮	⋮	⋮
⋮	⋮	⋮

図4 教師データイメージ

4.1 変換システムの流れ

本稿における、送り手と受け手のインタフェースとなるネガティブ言語変換システムの画面デザインを以下に示

す(図5)。デザインに関しては、受け手の趣味嗜好によって変更を行うことが理想であると考えているため、いくつかのパターンを用意した。以下に示すデザイン例は、その中の一例である。

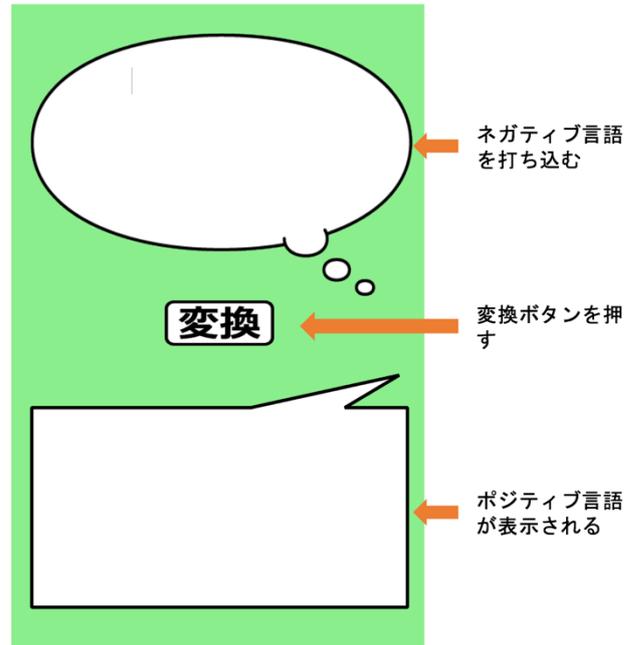


図5 変換画面デザイン例

受け手は、上部の吹き出しの中にネガティブ言語を入力する。次に変換ボタンを押すことで下部の吹き出しに、変換されたポジティブ言語が表示される。

4.2 ネガティブ言語変換例

本稿は言語変換の指針として、リフレーミングの技法並びにユーモアの考え方を技術思想としている。本システムで用いたポジティブ言語変換例を一部以下に述べる。なお、nega:はネガティブ言語、posi:はポジティブ言語変換例として表記する。

- (1) 送り手のネガティブ言語
「電車で食事をしていたら注意された。すごいムカつくし、ほんとストレスたまる。注意した奴、不幸になってほしい。」
- (2) 受け手のネガティブ言語変換例
(nega:)お前が非常識すぎる
(posi:)君の型にはまらない生き方はかっこいいよね!
初めて作ったたこ焼きを思い出したよ!

5. 学習精度の向上

4.1 手続き

機械学習は、教師データの質や量によって精度が変化するため、本システムにおいても学習精度は重要な指標で

ある。本研究においては、友人もしくは恋人からのネガティブ言語に限定した場面特異的な状況を想定しているため、当該場面に限定したネガティブ言語を画面上にランダム呈示し、素朴な態度を持つテストユーザーにネガティブ言語を受け取った後の素直な反応を入力してもらう。その後、システム内にて入力されたデータを教師データと照合し、最適に変換された回答を表示させ変換後の結果をテストユーザーにどの程度ポジティブに変換されたと感じたか、どの程度変換後の表現に納得度があるか、どの程度ユーモアを感じたかなどを全く思わないからとても思うまでの4件法にて評価してもらう。評価結果に応じて再度データをクローリング、教師データ増加のサイクルを繰り返しながら学習させることで変換精度を向上させていく。この際、同一のネガティブな文章が回数を重ねるごとにどのように変換が行われたかもあわせて記録し、変換精度の推移を含め、精度の変化を観察する(図6)。

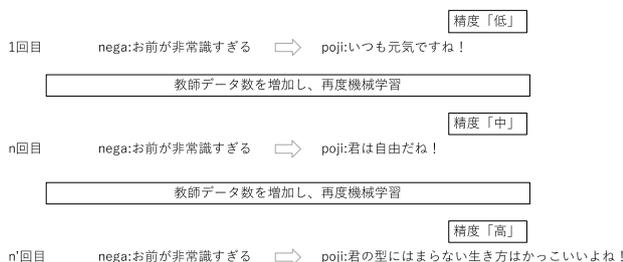


図6 学習精度向上の手続き

6. 今後の展望

今後の展望として、今回状況特異的な場面に限定して実装したシステムをより多様なコミュニケーションにも活用できるように汎用的な日本語コミュニケーションにおけるシステムにアップデートしていく必要があると考えている。

また、コミュニケーションを「楽しく」する支援ツールの役割として、その対極にあるコミュニケーション拒否を回避するだけでなく、心の負担を軽減しながら、より積極的なコミュニケーションの促進を支援できるシステムとして Entertainment Computing 技術をアップデートしていきたいと考えている。コミュニケーションには返報性があり、拒否的なコミュニケーションは悪循環となるリスクを生じさせるが、コミュニケーションをより前向きに「楽しそうだ」と感じてもらえる支援が実現すれば、嫌なヤツだと思ってコミュニケーションを避けていた相手が話してみたらいいヤツだったというような可能性を開くシステムとしてアップデートしていきたいと考えている。

また、いちいち文字入力で打ち込むのが面倒だというユーザーの要望にも音声入力機能等の実装によって応えていければと考えている。

本システムの利用によって相手の反応がマイルドに変化した過程を後になって見返すことができるような機能も実

装できればより「楽しく」コミュニケーションできるのではないかと考えている。

売り言葉に買い言葉という日本語があるように、相手に本当に言いたい言葉と違う言葉で返してしまうということもある。こうした緊張関係を緩和でき、本音で語り合える関係を作ることの支援もできるような、感情を抑える心の負担を人工知能の力で軽減し、気楽に受け入れられる、そんな Entertainment Computing を実現できたらと考えている。

謝辞

本研究に対してソーシャルヘルス研究会の先生方より貴重なご指導をいただいた。また、本研究に多大なるご理解とご協力をいただいた株式会社須永総合研究所並びに株式会社 SHP の先生方や同僚並びに関係者の皆様にこの場を借りて心から御礼を申し上げます。

参考文献

- 1) LINE, [https://line. me/ja/](https://line.me/ja/)
- 2) Twitter, [http://twitter. com/](http://twitter.com/)
- 3) 大坊郁夫:幸福を目指す対人社会心理学, ナカニシヤ出版, 2012.
- 4) レオン・フェスティンガー:認知的不協和の理論 社会心理学序説, 末永俊郎監訳, 誠信書房, 1965.
- 5) Adam D. I. Kramer: The spread of emotion via facebook, CHI 2012, ACM Press, pp, 767-770, 2012.
- 6) 大家眸美, 宮下芳明:ウェブコンテンツにおけるネガティブ感情表現の緩和手法, 情報処理学会研究報告, 2012, Vol, 2012-HCI-149 No. 12.
- 7) ネガボ辞典, <https://itunes.apple.com/jp/app/ネガボ辞典/id443101381?mt=8>
- 8) Romero, E. J., & Cruthirds, K. W. (2006). The use of humor in the workplace. *Academy of Management Perspectives*, 20, 58-69.
- 9) Holmes, J. (2006). Sharing a laugh: Pragmatic aspects of humor and gender in the workplace. *Journal of Pragmatics*, 38, 26-50.
- 10) Vinton, K. L. (1989). Humor in the workplace: It is more than telling jokes. *Small Group Behavior*, 20, 151-166
- 11) 上野行良 (2003). ユーモア現象に関する諸研究とユーモアの分類化について, *社会心理学研究*, 7, 112-120.
- 12) 宮戸美樹, 上野行良(1996) ユーモアの支援的効果の検討-支援的ユーモア志向尺度の構成, *心理学研究*, 67, 270-277.
- 13) 塚脇涼太, 樋口匡貴, 深田博巳:ユーモア表出と自己受容, 攻撃性, 愛他性との関係, *心理学研究*, 2009, 80, 4, pp.339-344.
- 14) 上野行良:ユーモアの心理学, サイエンス社, 2003
- 15) 牧野幸志:ユーモアと説得 深田博巳編著 説得心理学ハンドブック-説得コミュニケーション研究の最前線 2002 北大路書房
- 16) Dworkin, E. S. & Efran, J. S. (1967) The angered: Their susceptibility to varieties of humor. *Jornal of Personality and Social Psychology*, 6, 233-236.