

[穴埋め式]世界ことわざ辞典 /
[MASK]ed World Proverb Dictionary

「異なる文化で生まれた世界認識を、日本語のみを学習したモデルで捉え直す」

[穴埋め式]世界ことわざ辞典 / [MASK]ed World Proverb Dictionary は、日本語データのみから作られた言語モデルを使って、海外のことわざを再構成していくことでつくる、新しい辞典です。

世界の様々な国からことわざを収集、文中における単語をいくつかマスク=黒塗りしたうえで、言語モデルにその黒塗り箇所に入る単語を推測させます。

猿も■から落ちる

推測結果：木、屋根、椅子…

各ことわざについて、単語を 100 個推測させることで新たなテキストを生成。それらが原文と意味の近い順に掲載されています。

各ページを上から下へ順に読んでいくことで、元の意味からどんどんと離れていく、日本語データ独自の解釈による「ことわざ」を見ることができるでしょう。

機械学習における学習データの持つバイアスについて議論が生まれつつある昨今、本辞書ではあえて日本語のみを学習したモデルで、異文化の世界認識を表すテキストとも言えることわざの再構成を行っていきます。そこから日本語を扱う社会について新しい視点を与えられることができたら幸いです。

言語モデルについて：

言語モデルは、朝日新聞社の保有する記事から 2003 年以降に蓄積されたものを学習することで作られた、**Masked Language Model (MLM)** です。

MLM は文章中のマスクされた箇所にふさわしい単語を前後の文脈から予測することができ、今回はこちらのモデルを用いてことわざの生成を行なっています。

今日の天気は[mask]です。

今日の天気は[晴れ]です。

今日の天気は[曇り]です。

今日の天気は[順調]です。

今日の天気は[不安定]です。

今日の天気は[苦手]です。

…

参照元とライセンスについて：

今回掲載することわざは日本語版ウィキquote のカテゴリ：諺（下記リンク先）から抽出したものを、本作品におけるテキストはクリエイティブ・コモンズ (CC BY-SA 3.0) のもとで利用が可能です。

<https://ja.wikiquote.org/wiki/> カテゴリ：諺

ページの見方：

マスクされたことわざ
原文（国名）

生成されたことわざ

原文との意味の近さを表すスコア
(0~1の値。大きい方が意味が近い。)

マスクされたことわざ	原文（国名）	生成されたことわざ	原文との意味の近さを表すスコア
猿も■から落ちる	猿も木から落ちる (日本)	猿も木から落ちる	0.95
		猿も屋根から落ちる	0.85
		猿も椅子から落ちる	0.75
		猿も山から落ちる	0.65
		猿も川から落ちる	0.55
		猿も海から落ちる	0.45
		猿も空から落ちる	0.35
		猿も地から落ちる	0.25
		猿も火から落ちる	0.15
		猿も水から落ちる	0.05

意志が□の半分を引っ張る。
意志が重荷の半分を引っ張る。(アイスランド)

意志が自身の半分を引っ張る。	0.96301371901157
意志が彼の半分を引っ張る。	0.96146749871287
意志が自分の半分を引っ張る。	0.96146633443772
意志が思いの半分を引っ張る。	0.95933533983981
意志が僕の半分を引っ張る。	0.95845979885071
意志が私の半分を引っ張る。	0.95824767360072
意志が彼女の半分を引っ張る。	0.9566039278247
意志が努力の半分を引っ張る。	0.95528376855632
意志が人間の半分を引っ張る。	0.95432670085912
意志が言葉の半分を引っ張る。	0.95401072870828
意志が仕事の半分を引っ張る。	0.95352048470458
意志が己の半分を引っ張る。	0.95291884348985
意志が力の半分を引っ張る。	0.95226026898373
意志が気持ちの半分を引っ張る。	0.9521515334871
意志が人生の半分を引っ張る。	0.95179364716157
意志が家族の半分を引っ張る。	0.95124220102503
意志が俺の半分を引っ張る。	0.95110923692942
意志が信念の半分を引っ張る。	0.95106817885579
意志が期待の半分を引っ張る。	0.95088683474329
意志が相手の半分を引っ張る。	0.95048093349446
意志が話の半分を引っ張る。	0.94991901732717
意志が男の半分を引っ張る。	0.94877209339651
意志が経験の半分を引っ張る。	0.94808790752889
意志が息子の半分を引っ張る。	0.94805424401646
意志が体の半分を引っ張る。	0.94785860476526
意志が感情の半分を引っ張る。	0.94746475195528
意志が夢の半分を引っ張る。	0.94716721650499
意志が娘の半分を引っ張る。	0.94596738629878
意志が兄の半分を引っ張る。	0.94587407171436
意志が喜びの半分を引っ張る。	0.94563692348844
意志が成長の半分を引っ張る。	0.94516578981246
意志がメンバーの半分を引っ張る。	0.9444368839643
意志が頭の半分を引っ張る。	0.94441117438598
意志が希望の半分を引っ張る。	0.94424934628972
意志がチームの半分を引っ張る。	0.94416318604651
意志が胸の半分を引っ張る。	0.94404980844752
意志が生活の半分を引っ張る。	0.94400681893306
意志がライバルの半分を引っ張る。	0.94349283662576
意志が足の半分を引っ張る。	0.9433067625186
意志が周囲の半分を引っ張る。	0.94310557521398
意志が議論の半分を引っ張る。	0.94282066054667
意志が権力の半分を引っ張る。	0.94275605396457
意志が兄弟の半分を引っ張る。	0.94251722662227
意志が本業の半分を引っ張る。	0.94241253717609
意志が腹の半分を引っ張る。	0.9422721625242
意志が仲間の半分を引っ張る。	0.94218337447215
意志が体力の半分を引っ張る。	0.94157583239583
意志が実力の半分を引っ張る。	0.9398846565138
意志が活動の半分を引っ張る。	0.93985413585271
意志が費用の半分を引っ張る。	0.93984320948601
意志が映画の半分を引っ張る。	0.93982997613911
意志が元気の半分を引っ張る。	0.93928572794149
意志が応援の半分を引っ張る。	0.93897503264834
意志が代表の半分を引っ張る。	0.93807782781363
意志がパワーの半分を引っ張る。	0.93787344964136
意志がファンの半分を引っ張る。	0.93786571605971
意志がスタッフの半分を引っ張る。	0.93757655282588
意志が若手の半分を引っ張る。	0.93751110237975
意志が主力の半分を引っ張る。	0.93736119697869
意志が物語の半分を引っ張る。	0.93713852861966
意志が数字の半分を引っ張る。	0.93630313155691
意志が電話の半分を引っ張る。	0.93626037795699
意志が練習の半分を引っ張る。	0.93605707854896
意志が構想の半分を引っ張る。	0.93602275168675
意志が年齢の半分を引っ張る。	0.93581666829032
意志が取材の半分を引っ張る。	0.93423992408511
意志が医師の半分を引っ張る。	0.93419870133775
意志が選手の半分を引っ張る。	0.93391845823392
意志がアスリーの半分を引っ張る。	0.93347992518322
意志が現役の半分を引っ張る。	0.93347886417792
意志が攻撃の半分を引っ張る。	0.93318993849874
意志が収入の半分を引っ張る。	0.93298051369823
意志がテレビの半分を引っ張る。	0.93285934885854
意志が観客の半分を引っ張る。	0.93273810098593
意志が組織の半分を引っ張る。	0.93271166561577
意志が運動の半分を引っ張る。	0.93266714242468
意志が人気の半分を引っ張る。	0.9326305048038
意志が監督の半分を引っ張る。	0.93256125460752
意志がエースの半分を引っ張る。	0.93238873881098
意志が市場の半分を引っ張る。	0.93214349391536
意志が競技の半分を引っ張る。	0.93157514131848
意志が部員の半分を引っ張る。	0.93125361702026
意志が首相の半分を引っ張る。	0.93092261661694
意志がバンドの半分を引っ張る。	0.93078130820916
意志が率の半分を引っ張る。	0.92990691468619
意志が記録の半分を引っ張る。	0.92972786496026
意志が女子の半分を引っ張る。	0.92918512145573
意志が予算の半分を引っ張る。	0.92906787806821
意志が松井の半分を引っ張る。	0.92900157171727
意志が武士の半分を引っ張る。	0.92896443690918
意志がキャリアの半分を引っ張る。	0.92893528163146
意志がバットの半分を引っ張る。	0.9288222254461
意志が観衆の半分を引っ張る。	0.92859943864103
意志が野球の半分を引っ張る。	0.92817152143039
意志が体重の半分を引っ張る。	0.92807523271465
意志が試合の半分を引っ張る。	0.92799575374018
意志がサッカーの半分を引っ張る。	0.92733865581157
意志が幅の半分を引っ張る。	0.92693710475622
意志が村の半分を引っ張る。	0.92679650026594
意志が髪の毛の半分を引っ張る。	0.92612979991164

美しいものが美しいのではない、なものが美しいのである。
美しいものが美しいのではない、好きなものが美しいのである。(イタリア)

美しいものが美しいのではない、だめなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、ダメなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、すてきなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、もっともなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、きれいなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、しあわせなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、まれなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、たしかなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、ただなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、おしゃれなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、バカなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、ましなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、まじめなものが美しいのである。	0.99694109409826
美しいものが美しいのではない、結構なものが美しいのである。	0.99584053571428
美しいものが美しいのではない、正直なものが美しいのである。	0.99581669876205
美しいものが美しいのではない、嫌なものが美しいのである。	0.99565458260774
美しいものが美しいのではない、普通なものが美しいのである。	0.9955436854234
美しいものが美しいのではない、同じなものが美しいのである。	0.99547715042649
美しいものが美しいのではない、いやなものが美しいのである。	0.99547055695691
美しいものが美しいのではない、駄目なものが美しいのである。	0.99519246873895
美しいものが美しいのではない、地味なものが美しいのである。	0.99508238695071
美しいものが美しいのではない、同様なものが美しいのである。	0.99508059027484
美しいものが美しいのではない、別なものが美しいのである。	0.99501528873094
美しいものが美しいのではない、得意なものが美しいのである。	0.99482878805898
美しいものが美しいのではない、独特なものが美しいのである。	0.9943611291189
美しいものが美しいのではない、満足なものが美しいのである。	0.99430715465006
美しいものが美しいのではない、大事なものが美しいのである。	0.99424374525061
美しいものが美しいのではない、そっくりなものが美しいのである。	0.99413988136533
美しいものが美しいのではない、肝心なものが美しいのである。	0.99404491729172
美しいものが美しいのではない、純粋なものが美しいのである。	0.99400280798281
美しいものが美しいのではない、邪魔なものが美しいのである。	0.9938564097492
美しいものが美しいのではない、幸せなものが美しいのである。	0.99384583525804
美しいものが美しいのではない、余計なものが美しいのである。	0.99378118387777
美しいものが美しいのではない、完璧なものが美しいのである。	0.99370895559508
美しいものが美しいのではない、リアルなものが美しいのである。	0.99358814305669
美しいものが美しいのではない、大切なものが美しいのである。	0.99338845319884
美しいものが美しいのではない、味なものが美しいのである。	0.99322037861957
美しいものが美しいのではない、面倒なものが美しいのである。	0.99310138150208
美しいものが美しいのではない、粋なものが美しいのである。	0.99309764667323
美しいものが美しいのではない、ワイルドなものが美しいのである。	0.9930631249217
美しいものが美しいのではない、心配なものが美しいのである。	0.99291534615677
美しいものが美しいのではない、特別なものが美しいのである。	0.99281593132407
美しいものが美しいのではない、立派なものが美しいのである。	0.99266671186936
美しいものが美しいのではない、完全なものが美しいのである。	0.99258086477365
美しいものが美しいのではない、ベストなものが美しいのである。	0.99253339405777
美しいものが美しいのではない、ハッピーなものが美しいのである。	0.99243237514569
美しいものが美しいのではない、顕著なものが美しいのである。	0.99223723422006
美しいものが美しいのではない、公式なものが美しいのである。	0.992087039202
美しいものが美しいのではない、苦痛なものが美しいのである。	0.99207538660808
美しいものが美しいのではない、クールなものが美しいのである。	0.99204892332064
美しいものが美しいのではない、最悪なものが美しいのである。	0.99193704330687
美しいものが美しいのではない、上品なものが美しいのである。	0.99184692620614
美しいものが美しいのではない、ハードなものが美しいのである。	0.99182911273343
美しいものが美しいのではない、貧乏なものが美しいのである。	0.99162865681274
美しいものが美しいのではない、皮肉なものが美しいのである。	0.9916021170502
美しいものが美しいのではない、イヤなものが美しいのである。	0.99157853190268
美しいものが美しいのではない、美なものが美しいのである。	0.99155246218298
美しいものが美しいのではない、最低なものが美しいのである。	0.99153034678215
美しいものが美しいのではない、便利なものが美しいのである。	0.99142843521209
美しいものが美しいのではない、不幸なものが美しいのである。	0.99142763153326
美しいものが美しいのではない、ビッグなものが美しいのである。	0.99128656677697
美しいものが美しいのではない、不可欠なものが美しいのである。	0.99127839895751
美しいものが美しいのではない、優雅なものが美しいのである。	0.99125403764873
美しいものが美しいのではない、真っ黒なものが美しいのである。	0.99121360684117
美しいものが美しいのではない、残酷なものが美しいのである。	0.99120770440608
美しいものが美しいのではない、ホットなものが美しいのである。	0.9910590082851
美しいものが美しいのではない、ブラックなものが美しいのである。	0.99105228224324
美しいものが美しいのではない、平凡なものが美しいのである。	0.99101668742621
美しいものが美しいのではない、プライベートなものが美しいのである。	0.99094870532845
美しいものが美しいのではない、アートなものが美しいのである。	0.99072609187856
美しいものが美しいのではない、オープンなものが美しいのである。	0.99056797088919
美しいものが美しいのではない、純なものが美しいのである。	0.99051183376655
美しいものが美しいのではない、楽なものが美しいのである。	0.99049397919431
美しいものが美しいのではない、美味なものが美しいのである。	0.99045826696521
美しいものが美しいのではない、高級なものが美しいのである。	0.99043859777196
美しいものが美しいのではない、親切なものが美しいのである。	0.99033830539245
美しいものが美しいのではない、グレーなものが美しいのである。	0.99031070291087
美しいものが美しいのではない、クリアなものが美しいのである。	0.99025347658732
美しいものが美しいのではない、ソフトなものが美しいのである。	0.99022308348525
美しいものが美しいのではない、安全なものが美しいのである。	0.9902201540514
美しいものが美しいのではない、乙なものが美しいのである。	0.99011791595212
美しいものが美しいのではない、モダンなものが美しいのである。	0.99001303198312
美しいものが美しいのではない、幸運なものが美しいのである。	0.99000844451574
美しいものが美しいのではない、陽気なものが美しいのである。	0.98998910299763
美しいものが美しいのではない、幸福なものが美しいのである。	0.98991789653404
美しいものが美しいのではない、ローカルなものが美しいのである。	0.98991099936499
美しいものが美しいのではない、世界一なものが美しいのである。	0.9897564443013
美しいものが美しいのではない、エコなものが美しいのである。	0.98974681853762
美しいものが美しいのではない、希少なものが美しいのである。	0.98967809969038
美しいものが美しいのではない、スリムなものが美しいのである。	0.98960911602384
美しいものが美しいのではない、貴なものが美しいのである。	0.98960064371298
美しいものが美しいのではない、フェアなものが美しいのである。	0.9895554192133
美しいものが美しいのではない、簡素なものが美しいのである。	0.98951159705129
美しいものが美しいのではない、優美なものが美しいのである。	0.9895053412793
美しいものが美しいのではない、丈夫なものが美しいのである。	0.98948976443336
美しいものが美しいのではない、固有なものが美しいのである。	0.98933910513513
美しいものが美しいのではない、厳格なものが美しいのである。	0.98932388625789
美しいものが美しいのではない、不便なものが美しいのである。	0.9891915401506
美しいものが美しいのではない、平らなものが美しいのである。	0.98914222027518
美しいものが美しいのではない、快適なものが美しいのである。	0.98906984431136

いうこととすることの間には、**□**が横たわっている。
いうこととすることの間には、海が横たわっている。(イタリア)

いうこととすることの間には、ねじれが横たわっている。	0.99030726517435
いうこととすることの間には、いまが横たわっている。	0.99030726517435
いうこととすることの間には、なぞが横たわっている。	0.99030726517435
いうこととすることの間には、そこが横たわっている。	0.99030726517435
いうこととすることの間には、なにが横たわっている。	0.99030726517435
いうこととすることの間には、それぞれが横たわっている。	0.99030726517435
いうこととすることの間には、いろいろが横たわっている。	0.99030726517435
いうこととすることの間には、かわりが横たわっている。	0.99030726517435
いうこととすることの間には、だれもが横たわっている。	0.98960195929422
いうこととすることの間には、今が横たわっている。	0.98895135287477
いうこととすることの間には、別が横たわっている。	0.98700216895005
いうこととすることの間には、違うが横たわっている。	0.98687597893769
いうこととすることの間には、いくつか横たわっている。	0.98668710262541
いうこととすることの間には、物語が横たわっている。	0.98657831327875
いうこととすることの間には、苦しみが横たわっている。	0.98626677960403
いうこととすることの間には、悲しみが横たわっている。	0.98626677960403
いうこととすることの間には、理由が横たわっている。	0.98601807326043
いうこととすることの間には、テーマが横たわっている。	0.98592810946924
いうこととすることの間には、何かが横たわっている。	0.98575933881768
いうこととすることの間には、危険が横たわっている。	0.98553069119593
いうこととすることの間には、謎が横たわっている。	0.98529001985906
いうこととすることの間には、無理が横たわっている。	0.98528018750482
いうこととすることの間には、違いが横たわっている。	0.9852763779583
いうこととすることの間には、現実が横たわっている。	0.98514255446211
いうこととすることの間には、偶然が横たわっている。	0.98434694912376
いうこととすることの間には、問題が横たわっている。	0.98427081771162
いうこととすることの間には、この問題が横たわっている。	0.9842707679864
いうこととすることの間には、背景が横たわっている。	0.98412996377503
いうこととすることの間には、壁が横たわっている。	0.98397956023858
いうこととすることの間には、未知が横たわっている。	0.98392112839427
いうこととすることの間には、変化が横たわっている。	0.98374243682537
いうこととすることの間には、歴史が横たわっている。	0.98362397825443
いうこととすることの間には、その理由が横たわっている。	0.98361872543537
いうこととすることの間には、ヒントが横たわっている。	0.98319539202678
いうこととすることの間には、恐怖が横たわっている。	0.9830435101579
いうこととすることの間には、悲劇が横たわっている。	0.98286089022535
いうこととすることの間には、不安が横たわっている。	0.98274666940813
いうこととすることの間には、答えが横たわっている。	0.98274553000062
いうこととすることの間には、失敗が横たわっている。	0.98272310766152
いうこととすることの間には、疑問が横たわっている。	0.98259292605215
いうこととすることの間には、赤ちゃんが横たわっている。	0.98258624090392
いうこととすることの間には、困難が横たわっている。	0.98247835471548
いうこととすることの間には、闇が横たわっている。	0.9824097056394
いうこととすることの間には、戦争が横たわっている。	0.98238974683097
いうこととすることの間には、孤独が横たわっている。	0.9822937160662
いうこととすることの間には、本質が横たわっている。	0.98217282305688
いうこととすることの間には、扉が横たわっている。	0.98188082063995
いうこととすることの間には、誤解が横たわっている。	0.98183441818729
いうこととすることの間には、苦悩が横たわっている。	0.98170314127113
いうこととすることの間には、限界が横たわっている。	0.98168325383888
いうこととすることの間には、模索が横たわっている。	0.98168151351518
いうこととすることの間には、危機が横たわっている。	0.98167090278719
いうこととすることの間には、矛盾が横たわっている。	0.98163577867428
いうこととすることの間には、課題が横たわっている。	0.98162653528578
いうこととすることの間には、事情が横たわっている。	0.9815970277752
いうこととすることの間には、迷いが横たわっている。	0.98158817189607
いうこととすることの間には、複雑が横たわっている。	0.98157136926421
いうこととすることの間には、要因が横たわっている。	0.98146501522129
いうこととすることの間には、穴が横たわっている。	0.98109879816342
いうこととすることの間には、混乱が横たわっている。	0.98106876008936
いうこととすることの間には、感情が横たわっている。	0.9807789722856
いうこととすることの間には、懸念が横たわっている。	0.98076065132062
いうこととすることの間には、ドラマが横たわっている。	0.98074639321657
いうこととすることの間には、国境が横たわっている。	0.98067295646367
いうこととすることの間には、構図が横たわっている。	0.98055792718104
いうこととすることの間には、ルールが横たわっている。	0.98054390327015
いうこととすることの間には、不満が横たわっている。	0.9804751169274
いうこととすることの間には、異論が横たわっている。	0.98044844974718
いうこととすることの間には、ストレスが横たわっている。	0.98028023406236
いうこととすることの間には、バランスが横たわっている。	0.98027974724332
いうこととすることの間には、疑念が横たわっている。	0.98026739967842
いうこととすることの間には、思惑が横たわっている。	0.98016244320849
いうこととすることの間には、信念が横たわっている。	0.98013164177586
いうこととすることの間には、構造が横たわっている。	0.98013001853887
いうこととすることの間には、葛藤が横たわっている。	0.9800774097644
いうこととすることの間には、二つが横たわっている。	0.98001863623632
いうこととすることの間には、国家が横たわっている。	0.97999872200293
いうこととすることの間には、根が横たわっている。	0.97998329359764
いうこととすることの間には、隔たりが横たわっている。	0.97992621485765
いうこととすることの間には、緊張が横たわっている。	0.97988107082022
いうこととすることの間には、差が横たわっている。	0.97978128512935
いうこととすることの間には、リスクが横たわっている。	0.97957620484576
いうこととすることの間には、境界が横たわっている。	0.97954662867806
いうこととすることの間には、基礎が横たわっている。	0.97925186071836
いうこととすることの間には、距離が横たわっている。	0.97917958643684
いうこととすることの間には、次元が横たわっている。	0.9791363239389
いうこととすることの間には、ハードルが横たわっている。	0.97902933551561
いうこととすることの間には、鍵が横たわっている。	0.97901074855988
いうこととすることの間には、分断が横たわっている。	0.97882124125594
いうこととすることの間には、一つが横たわっている。	0.97881835795274
いうこととすることの間には、出口が横たわっている。	0.97869563726746
いうこととすることの間には、対立が横たわっている。	0.97864162987366
いうこととすることの間には、トラブルが横たわっている。	0.97837298582035
いうこととすることの間には、論争が横たわっている。	0.97833915455774
いうこととすることの間には、偏見が横たわっている。	0.97822105849485
いうこととすることの間には、閥門が横たわっている。	0.97819109419746
いうこととすることの間には、哲学が横たわっている。	0.9780495620948
いうこととすることの間には、責任が横たわっている。	0.97800795325874
いうこととすることの間には、疑いが横たわっている。	0.9779695314709
いうこととすることの間には、真理が横たわっている。	0.97784504613411

