

# OKR における表現の具体度が Objective と Key Results の整合性に与える影響

## Impact of Expression Concreteness in OKRs on the Alignment Between Objectives and Key Results

本江 拓海<sup>1</sup>, 江刺 裕太<sup>1</sup>, 松木 萌<sup>1</sup>, 大森 隆司<sup>1</sup>, 大澤 正彦<sup>1</sup>,  
Takumi Hongo<sup>1</sup>, Yuta Esashi<sup>1</sup>, Moe Matsuki<sup>1</sup>, Takashi Omori<sup>1</sup>, Masahiko Osawa<sup>1</sup>

<sup>1</sup> 日本大学 文理学部 次世代社会研究センター (RINGS)

<sup>1</sup> Research Institute for Next Generation Society,  
College of Humanities and Sciences, Nihon University

**Abstract:** 目標管理法 OKR は、全ての Key Results (KR) を達成することで Objective (O) が達成されるという整合性が重要だが、OKR をどのように改善すれば整合性が高まるかを客観的に知るべきでない。本研究では「O と KR の整合性は、OKR に含まれる各単語を具体化することで高くなり、抽象化することで低くなる」という仮説をたて、23 件の専門家が設定した OKR を用いて検証を行った。結果として、KR のみの具体化および O と KR の同時具体化はいずれも整合性スコアを有意に向上させた。一方、抽象化については整合性が一律に低下する傾向は確認されず、条件によっては向上するケースも観察された。整合性を左右する本質は単語の具体度ではなく、単語変換が O の想定する文脈の内部に留まり、その文脈を「深掘り／言い換え」できているかという文脈の一貫性にあると考察した。

## 1 はじめに

OKR (Objectives and Key Results) は、目標 (Objective) と主要な成果指標 (Key Results) をセットで管理するフレームワークであり、近年、個人・チーム・組織の目標設定に広く採用されている。OKR はシリコンバレーの企業で発展し、Google の導入を機に広く知られるようになった。現在では成長戦略やチーム強化の手法として多くの企業が採用している。OKR は、定性的な目標である Objective に対し、2~5 個の定量的な成果目標である Key Results を整合させるのが基本であり、「すべての Key Results を達成することで Objective も達成される」ことが前提とされている [1]。しかし、定性的な Objective と定量的な Key Results を整合させることが実務上困難であるという課題が報告されている [2]。こうした背景を受けて、Objective と Key Results の整合性を担保するために、他の目標管理法との統合を試みる研究も存在する [3]。しかし、Objective と Key Results の整合性に影響を与える要因として、目標記述に用いられる言語表現そのものに着目した研究は十分に行われていない。本研究では「Objective と Key Results の整合性は、OKR に含まれる各単語を具体化

することで高くなり、抽象化することで低くなる」という仮説をたて、23 件の専門家が設定した OKR を用いて検証を行った。以降、Objective を O、Key Result を KR と表記する。

## 2 関連研究

### 2.1 Objectives and Key Results (OKR)

OKR は、組織や個人が達成すべき目標と、その達成度を評価するための主要な成果指標をセットで定義する目標管理のフレームワークである [1]。OKR は以下の 2 つの要素から構成される：

**Objective** 達成したい目標を定性的に示す。方向性や意図を端的に表現するもの。

**Key Results** Objective の達成度を定量的に評価するための主要な成果指標。

例えば、以下はコンサルティングサイト Workpath GmbH [6] で公開されている「E-Commerce OKRs」に基づく、カスタマーサポートの質向上を目指す OKR の一例である：

**Objective** Increase the on-time delivery rate for E-

commerce Logistics, ensuring customer satisfaction and trust in our services

Key Results1 Improve Package Tracking Accuracy: 95% of all tracking updates are correct

Key Results1 Improve Package Tracking Accuracy: 95% of all tracking updates are correct

Key Results2 Reduce Delivery Errors: Error rate below 2% for all deliveries

Key Results3 Boost Delivery Speed: Average delivery time under 24 hours

Key Results4 Enhance Customer Support: 90% positive feedback on support interactions

Key Results5 Optimize Routing Efficiency: 85% of routes use optimal paths

この例では、「E コマース物流における納期遵守率の向上」という定性的な目標 (O) に対して、追跡精度や配送エラー率、平均納品時間、顧客対応の評価、ルート最適化率などの測定可能な項目が KR として紐づけられている。O が「何を指すか」を示す一方で、KR はその達成状況を測る具体的な成果を提供している。すなわち、O は複数の KR によって意味的に構成される全体として位置付けられるため、O とそれぞれの KR は Part-of 関係であると言える。

## 2.2 O と KR の整合性に関する研究

O と KR を整合させることが困難であるという実務上の課題に対して、いくつかの試みがある。例えば、OKR を他の目標管理手法と統合する枠組みが提案されており、目標の階層構造や戦略との接続を明確にすることで整合性を担保しようとする試みがある [3]。しかし、これは目標管理プロセスや構造設計に着目したものであり、目標文に用いられる言語表現そのものが整合性に与える影響については十分に検討されていない。一方で、我々は埋め込み表現を用いて定量的に整合性を評価する手法を提案した [8]。この手法は「適切に設定された OKR では、O の埋め込み表現は対応する KR の埋め込み表現の総和と高い類似度を持つ」という仮説のもと、コサイン類似度によって整合性を数値化する手法である。この手法を専門家が設定した OKR で評価した結果、対応する KR 群と対応しない KR 群の間に統計的に有意な差が確認され、OKR の整合性を客観的に評価できる可能性が示された。このように、整合性を客観的に評価する指標はあるものの、整合性を向上させる支援はない。本研究は、整合性を向上させるアプローチとして、OKR の文章を具体化することに着目した。

## 2.3 具体化抽象化に関する研究

単語の具体性・抽象性は、人間の認知特性に関する指標であると同時に、言語表現としての意味構造にも深く関与する性質である。既存研究では、大規模な単語に対して「具体性」の評定値を付与し、各単語がどの程度感覚的・状況依存的な対象を指示するかを数量化したことが報告されている [9]。このような単語の具体性は、分布意味論の枠組みにおいても重要な役割を果たすことが示されている。

具体語と抽象語の分布的特徴を分析した研究では、両者が異なる文脈の振る舞いを示すことが報告されている。具体語は比較的限定された状況や対象に結びついて使用されるため、使用文脈が収束しやすく、埋め込み表現で密なクラスタを形成する傾向がある。一方、抽象語は多様な文脈で用いられるため、分布が広く拡散しやすく、意味的に多義的な位置を占める傾向がある [10]。この差異は、単語の具体性が、埋め込み表現における意味的配置や距離関係に影響を及ぼすことを示唆している。

さらに、人間の意味判断と大規模言語モデルの表現との整合性が、単語の具体性によって体系的に左右されることも示されている。具体語は抽象語に比べて、埋め込み表現空間において意味ベクトルの分散が小さくなり、表現間の対応関係が明確になりやすい [11]。

## 3 仮説

本研究では、OKR に含まれる単語を具体化することで、元の OKR よりも O と KR の整合性が高くなるという仮説を立てる。本章では本仮説を立てるに至った論拠を述べる。

### 3.1 OKR の整合性を定量的に評価する方法

我々は埋め込み表現を用いて定量的に整合性を評価する手法を提案した [8]。この手法は「適切に設定された OKR では、O の埋め込み表現は対応する KR の埋め込み表現の総和と高い類似度を持つ」という仮説のもと、コサイン類似度によって整合性を数値化する手法である。O の埋め込みベクトルを  $\mathbf{w}_O$ 、各 KR の埋め込みベクトルを  $\mathbf{w}_{KR_i}$  とすると、整合性スコアは以下で定義される。

$$Alignment(O, KR) = \cos \left( \mathbf{w}_O, \sum_{i=1}^n \mathbf{w}_{KR_i} \right)$$

ここでコサイン類似度は、

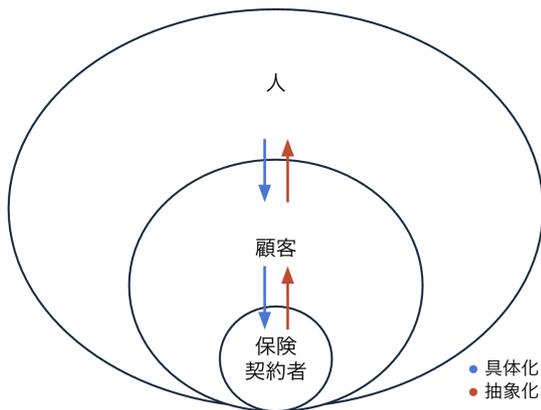


図1 具体化と抽象化の概念

$$\cos(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|}$$

で定義される。この値は  $-1$  から  $1$  の範囲を取り、 $1$  に近いほど両ベクトルは同方向であり、 $0$  と  $KR$  の意味的整合性が高いことを示す。

### 3.2 具体化することで整合性が高くなる理由

図1のように抽象化とは、より多くの事例を包含する上位概念への移動であり、具体化とは、適用範囲を限定する下位概念への移動である。例えば、「顧客」は「保険契約者」よりも広い意味領域を持つ上位概念であり、具体化は意味領域を縮小させ、抽象化は意味領域を拡張させる操作とみなせる。

また、意味領域を縮小させる具体化は、語が指示し得る状況や対象に制約を加えて適用範囲を限定する操作でもある。つまり、具体語は使用文脈が限定されるため、埋め込み表現における特定の意味成分が強調され、ベクトル方向がより収束すると考えられる。一方で、抽象語は複数の意味的側面を内包しやすく多様な文脈で用いられる。つまり、埋め込み表現におけるベクトル方向は収束しづらいと考えられる。

$O$  と  $KR$  の整合度は埋め込み表現間のコサイン類似度で量化されるため、単語を具体化することでその意味方向が明確になり、 $O$  と  $KR$  が共有する潜在的意味成分の割合が増加すれば、方向の一致度が高まり整合性の数値は上昇すると予測される。

## 4 検証方法

### 4.1 OKR データ

本実験では、コンサルティングサイト Workpath GmbH[6] で公開されている 23 件の OKR サンプルデータを用いた。Workpath GmbH は、専門的知見を有する企業向けコンサルティング会社であり、本研究では整合性の高い OKR データとして同サイトのデータを採用した。OKR は業種や生活の分野を想定して作成されたもので、多様な業界（金融、IT、ヘルスケア、非営利組織、スタートアップなど）や個人向け目標を幅広くカバーしている。

### 4.2 前処理方法

仮説の検証を行うために、OKR の自然言語を埋め込み表現に変換する。変換手段として、OpenAI が提供する text-embedding-ada-002 を使用する。このモデルは、任意の長さのテキストを 1536 次元の高次元ベクトルに変換する。この高次元のベクトルは意味的な情報を保持した埋め込み表現である。Transformer アーキテクチャに基づき、検索、分類、意味的類似度評価など多様な NLP タスクで活用されている [7]。

### 4.3 単語置換に基づく OKR 変換手法

本手法は、単語の抽象度に基づいて変換対象語を選択し、それをより具体的またはより抽象的な単語へと置換する。そうすることで、OKR の表現粒度を制御する。

まず、初期状態の OKR について、 $O$  および各  $KR$  を埋め込みベクトルに変換し、 $O$  ベクトルと  $KR$  ベクトルのコサイン類似度を算出する。これを変換前の整合性の値 ( $\cos_{\text{before}}$ ) とする。

次に、変換対象となる文に対して形態素解析を行い、トークン単位に分割する。各トークンには、既存の抽象度辞書 [5] を参照して抽象度スコアを付与する。助詞や記号など、抽象度が定義できない単語は対象から除外する。

抽象化条件では、対象文中で最も具体的な単語（抽象度スコアが最小の単語）を変換対象とし、具体化条件では、最も抽象的な単語（抽象度スコアが最大の単語）を変換対象とする。選択された単語に対して、辞書および候補生成規則に基づき、より具体的（あるいはより抽象的）な意味をもつ単語を候補として取得し、文中の当該単語を 1 箇所のみ置換することで変換後文を生成する。

生成した変換後 OKR について、再度文埋め込みを計算し、O ベクトルと KR ベクトルのコサイン類似度を算出する。これを変換後の整合性の値 (cos\_after) とする。

#### 4.4 条件

本研究では、上記の手法を適用する範囲の違いにより、次の 2 条件を設定する。

- 検証 1：KR のみに単語置換を適用し、O は固定する条件
- 検証 2：O および KR の両方に単語置換を適用する条件

検証 1 では、O を固定したまま KR のみの具体度を操作することで、KR 側の具体度変化が整合性に与える影響を測定する。OKR の設計上、KR は O よりも具体度が高い構造を持つ。KR をさらに具体化することで、O の意味領域との対応がより明確になり、整合性が向上する可能性があると考えた。

検証 2 では、O と KR の双方の具体度を同時に操作することで、OKR 全体の具体度が整合性に与える影響を検討する。具体度が低い目標は解釈の幅が広がる可能性があるため、両者を具体化することで解釈の揺れが抑えられ、「KR を達成すれば O が達成される」という構造がより明確になると考えた。

両条件は同一の置換手法に基づいており、単語置換の適用範囲のみが異なる。これにより、操作範囲の違いが整合性の値に与える効果を比較する。

## 5 検証結果

### 5.1 検証 1：KR のみに適用した結果

#### 5.1.1 具体化

OKR データに対して、置換前の OKR の整合性の値と置換後の OKR の整合性の値を算出し、箱ひげ図にしたものを図 2 に示す。具体的に変換前の中央値は 0.8909 で、変換後の中央値は 0.8974 であり、その差分は 0.0065 である。検定の結果、統計的に有意な差が認められた ( $p=0.000165$ )。この結果から変換後の中央値は変換前よりも上昇していることが確認できる。

#### 5.1.2 抽象化

同様に置換前と置換後の整合性の値を箱ひげ図にしたものを図 3 に示す。具体的に変換前の中央値は 0.8909 で、変換後の中央値は 0.8918 であり、その差分は 0.0009

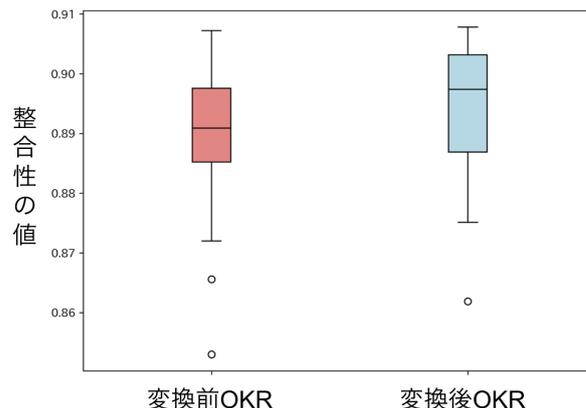


図 2 検証 1 の結果 (KR のみ具体化)

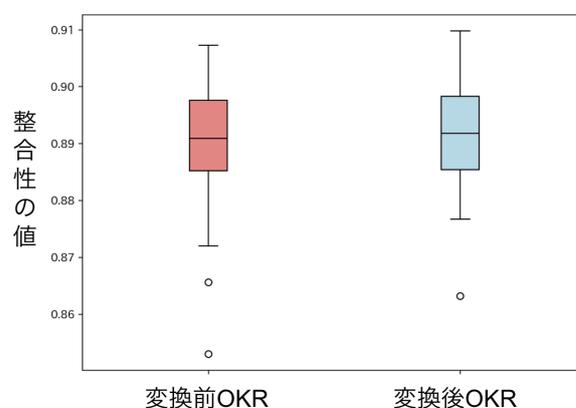


図 3 検証 1 の結果 (KR のみ抽象化)

である。検定の結果、統計的に有意な差が認められた ( $p=0.008189$ )。この結果から、置換前後の分布は大きく崩れておらず、中央値は同程度またはわずかに上昇していることが確認された。

### 5.2 検証 2：O と KR 両方の結果

#### 5.2.1 具体化

結果を図 4 に示す。具体的に変換前の中央値は 0.8909 で、変換後の中央値は 0.8968 であり、その差分は 0.0059 である。検定の結果、統計的に有意な差が認められた ( $p=0.039853$ )。この結果より、変換後の中央値は上昇していることがわかる。

#### 5.2.2 抽象化

図 5 に結果を示す。具体的に変換前の中央値は 0.8909 で、変換後の中央値は 0.8906 であり、その差分は -0.0003 である。検定の結果、統計的に有意な差が認められなかった ( $p=0.169523$ )。結果から、変換前後の分布は強

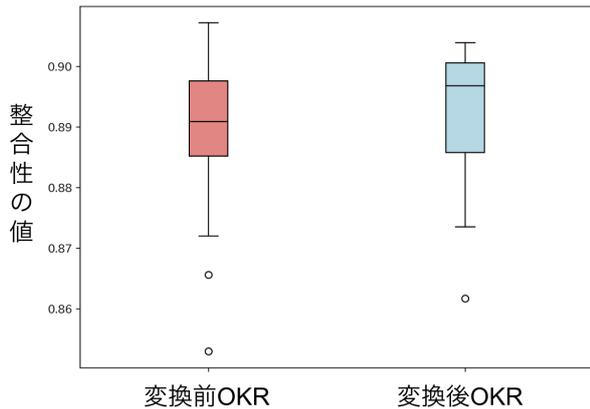


図4 検証2の結果(OとKR具体化)

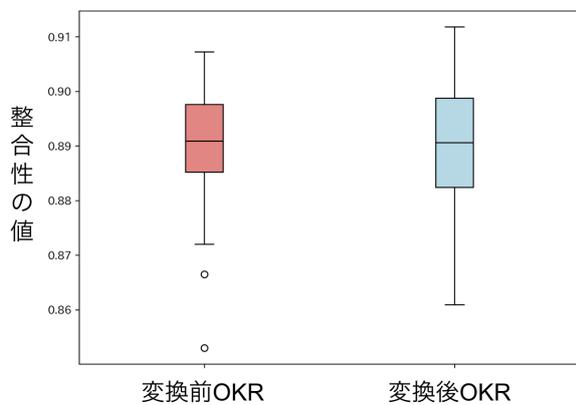


図5 検証2の結果(OとKR抽象化)

く重なっており、明確な一方向へのシフトは確認されない。

### 5.3 条件間の比較

図6に具体化における整合性の差( $\Delta$  整合性の値)の条件間比較を、図7に抽象化における $\Delta$  整合性の値の条件間比較をそれぞれ示す。ここで $\Delta$  整合性の値は、操作後のコサイン類似度から操作前のコサイン類似度を減じた値( $\Delta$  整合性の値 =  $\text{COS}_{\text{after}} - \text{COS}_{\text{before}}$ )であり、正の値は整合性の向上を、負の値は整合性の低下を示す。なお、ばらつきの指標として四分位範囲(interquartile range,  $\text{IQR} = \text{Q3} - \text{Q1}$ )を用いた。

まず具体化(図6)について見ると、KRのみの具体化では中央値は0.0036であり、IQRは0.00525であった( $\text{Q1}=0.00155$ ,  $\text{Q3}=0.0068$ )。最小値は-0.0033, 最大値は0.0182であり、多くのサンプルが整合性の向上を示している。

一方、OとKRを同時に具体化した条件では、中央値

は0.0030で、IQRは0.00895であった( $\text{Q1}=-0.00275$ ,  $\text{Q3}=0.0062$ )。最小値は-0.0072, 最大値は0.0169であり、KRのみの場合と比較して負方向への広がりが大きく、分布のばらつきが大きいことが確認された。

次に抽象化(図7)について見ると、KRのみの抽象化では中央値は0.0023, IQRは0.00350であった( $\text{Q1}=0.00065$ ,  $\text{Q3}=0.00415$ )。最小値は-0.0049, 最大値は0.0115であり、正負のサンプルが混在しているものの、中央値は正側に位置している。

さらに、OとKRを同時に抽象化した条件では、中央値は0.0026, IQRは0.00695であった( $\text{Q1}=-0.0013$ ,  $\text{Q3}=0.00565$ )。最小値は-0.0156, 最大値は0.011であり、負方向への広がりが顕著であった。

以上を総合すると、具体化は全体として整合性を押し上げる傾向があるが、特にKRのみの具体化が最も安定的である。一方、抽象化は一律に整合性を低下させる操作ではないものの、Oを同時に操作すると分散が拡大し、結果が不安定になる傾向が確認された。

これらの結果は、整合性の変化が単語の抽象度そのものよりも、操作対象(KRのみか、Oを含むか)によって大きく左右される可能性を示唆している。

## 6 考察

### 6.1 検証1: KRのみの操作に関する考察

KRのみを具体化・抽象化した結果を比較すると、OKRの整合性を左右する本質は「単語の具体度」そのものではなく、その変換が「Oが描く文脈の内部に収まっているかどうか」であると考えられる。

まず、具体化によって整合性が向上した例を述べると、customer → policyholderが挙げられる。この変換では、Oが想定する「保険」という文脈の内側で単語が細分化され、KRがOをより高い解像度で描写する文へと変化した。その結果、OとKRが共有する潜在的な意味が明確化され、意味的距離が縮まった。

一方で、整合性が低下した例として、project → enterpriseがある。この場合、KRの文脈が「現場」から「組織全体」へと横滑りし、Oとは異なる意味領域となっている。結果として、KRがOとは別の文脈を形成し、両者の対応関係が弱まった。

これらの結果から、具体化は単に適用範囲を限定する下位概念への移動として行われるのではなく、「Oの文脈を内側から掘り下げる方向」で行われたときにのみ有効であることが分かる。

一方で、抽象化によって整合性が低下した例として usage → practice, mobility → quality,

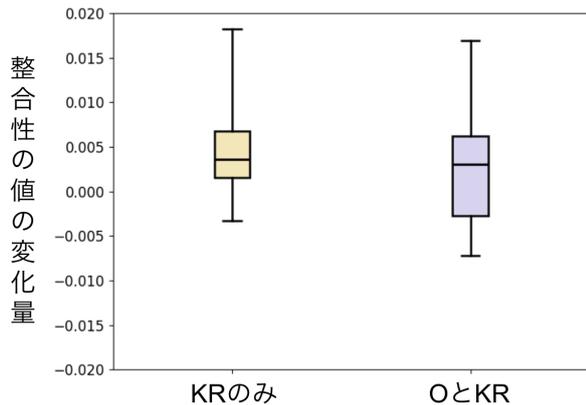


図6 KRのみ具体化とOとKRの具体化の比較

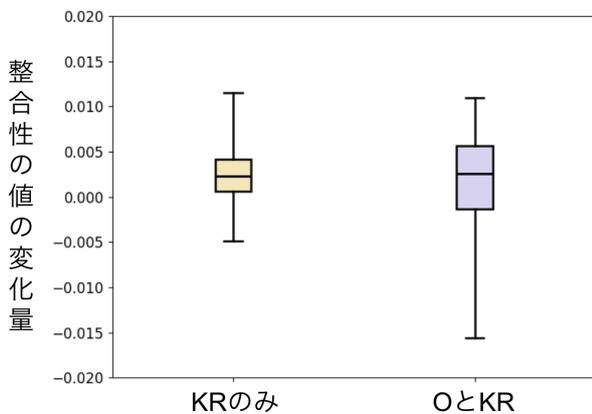


図7 KRのみ抽象化とOとKRの抽象化の比較

sustainability → property が挙げられる。これは本来「環境施策」や「持続可能性」を指していた単語が、「実践」「品質」「資産」といった汎用的で別軸の概念へと置き換えられた。その結果、「グリーンモビリティの推進」が「グリーン品質の促進」のように解釈され得る状態となり、抽象度は上がっているにもかかわらず、評価軸や主題がずれ、Oの前提文脈が失われたため、整合性が低下した。

次に、抽象化によって整合性が向上した例として、courses → activity, partnerships → human relationship, readiness → state of mind が挙げられる。ここでは、具体的な施策や行動を表す単語が、「活動」「関係性」「状態」といったOが目指す状態レベルの表現へと寄っている。これにより、KRは単なる数値目標ではなく、「Oの目標像を別の角度から言い換える文」に近づき、OとKRが同一の文脈を共有するようになった結果、意味的距離が縮まったと考えられる。

以上より、KRの具体化・抽象化は、その変換が「Oの文脈を内側から掘り下げる／言い換える方向」で行われる限り、OKRの整合性を高める。しかし、Oとは異

なる文脈や評価軸へKRをずらす変換は、具体・抽象を問わず、整合性を低下させる。

## 6.2 検証2: OとKR両方の操作に関する考察

OとKRを同時に具体化・抽象化した結果を比較すると、検証1の考察と同様に、OKRの整合性を左右する本質は「単語の具体度」そのものではなく、その変換が「Oが描く文脈の内部に収まっているかどうか」であることが分かった。

例えば、具体化によって整合性が向上した例として、insurance → whole life insurance, customer → policyholder が挙げられる。ここでは、Oが想定する「保険業務」という文脈の内側で、OとKRの双方が同じ方向に精緻化された。一方で、整合性が低下した例として、project → enterprise がある。この変換では、Oが「プロジェクト管理」という枠組みから、KRが「企業全体」という別文脈へと横滑りし、OとKRの文脈が乖離した。これらは検証1の考察と同様、KRがOをより高い解像度で描写する文への変換が起きることでの整合性が向上、文脈の横滑りが起きることによる整合性の低下が起きたと考えられる。

抽象化においても、整合性が低下した例として、sustainability → attribute, mobility → quality などの変換がある。ここでは、本来「環境」や「エネルギー」という領域を指していた単語が、「属性」「品質」といった別軸の概念へと移動し、Oの前提とする問題領域そのものが失われた。一方で、整合性が向上した例として、partnerships → human relationship, readiness → state of mind が挙げられる。これは、KRが「施策レベル」から「状態・姿勢レベル」へと引き上げられ、Oが描く「健康を大切にしている在り方」という文脈の内部に収束した。これらの例から、評価軸がずれることによる整合性の低下、文脈が別の角度から言い変わることで整合性の向上が起きたと考えられる。

以上より、OとKRを同時に具体化・抽象化した場合であっても、整合性を高める条件は、「その変換がOの文脈の内部で行われているか」に尽きる。

また検証1とは異なるOの変換による影響として、具体化は、Oの文脈を内側から精緻化する方向であれば整合性を高めるが、文脈を越境する具体化は整合性を低下させる。抽象化もまた、Oの文脈を保持したまま本質を言い換える場合にのみ有効であり、問題領域を曖昧化・移動させる抽象化は整合性を損なう。このようにKRだけでなくOの変換の影響があることで、変動幅が大きく、不安定化する要因となったと考えられる。

## 7 まとめ

本研究では、OKRにおけるOとKRの整合性に対して、単語の抽象度が与える影響を検証した。23件の専門家作成OKRを対象に、単語の具体化・抽象化を行い、埋め込み表現に基づくコサイン類似度によって整合性の変化を定量的に評価した。

その結果、KRのみの具体化は整合性を比較的安定的に向上させる傾向が確認された。一方で、抽象化は一律に整合性を低下させる操作ではなく、条件によっては向上する事例も観察された。また、OとKRを同時に操作する条件では変動幅が大きく、不安定化する傾向が見られた。

これらの結果から、整合性を左右する本質は単語の「具体／抽象」そのものではなく、単語変換がOの想定する文脈の内部に留まり、その文脈を深掘りあるいは言い換える方向で行われているかどうかにあることが示唆された。すなわち、OとKRが同一の文脈を共有したまま具体度のみが変化しているかという文脈の一貫性が重要である。

今後は、文全体の意味構造を考慮した変換手法の検討や、実務データへの適用を通じて、OKR設計支援への発展を目指す。

## 参考文献

- [1] Doerr, J.: Measure What Matters, Penguin Random House (2018)
- [2] Stray, V., Gundelsby, J. H., Ulfesnes, R., Moe, N. B.: How Agile Teams Make Objectives and Key Results (OKRs) Work, Proceedings of the International Conference on Software and System Processes and International Conference on Global Software Engineering (ICSSP '22), pp. 104-109, Association for Computing Machinery (2022)
- [3] Trinkenreich, B., Santos, G., Barcellos, M., Conte, T.: Combining GQM+Strategies and OKR: Preliminary Results from a Participative Case Study in Industry, (2019)
- [4] Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space, arXiv preprint arXiv:1301.3781 (2013). Available at: <https://arxiv.org/abs/1301.3781>
- [5] Butterfuss, R., Doran, H.: An Application of Text Embeddings to Support Alignment of Educational Content Standards, Educational Measurement: Issues and Practice, Vol. 44, No. 1, pp. 73-83 (2024). DOI: 10.1111/emip.12641
- [6] Workpath: OKR Examples, Retrieved from <https://www.workpath.com/en/okr-examples> (2024).
- [7] OpenAI: OpenAI Platform Documentation: Embeddings, Retrieved from <https://platform.openai.com/docs/guides/embeddings> (2023).
- [8] 江刺 裕太, 松木 萌, 三浦 紘嵩, 大澤 正彦: 埋め込み表現を活用したOKRのObjectiveとKey Resultsの定量的整合性分析, 2025年度日本認知科学会第42回大会予稿集, P3-52 (2025)
- [9] Brysbaert, M., Warriner, A. B., Kuperman, V.: Concreteness Ratings for 40,000 Generally Known English Word Lemmas, Behavior Research Methods, Vol. 46, No. 3, pp. 904-911 (2014). DOI: 10.3758/s13428-013-0403-5
- [10] Frassinelli, D., Schulte im Walde, S.: Distributional Properties of Concrete and Abstract Words in Contextualized Word Embeddings, Proceedings of the 13th Linguistic Annotation Workshop (LAW XIII), pp. 104-113 (2019)
- [11] Iaia, C., Choksi, B., Wiebers, E., Roig, G., Fiebach, C. J.: The Representational Alignment between Humans and Language Models is Implicitly Driven by a Concreteness Effect, arXiv preprint arXiv:2505.15682 (2025)