

# RT2text em Ação: Análise Semântica de Ocorrências de Segurança Pública em Curitiba

1

**Abstract.** *Trajectory summarization is a vital field in mobility analysis, enabling the extraction of key patterns from large datasets. This paper demonstrates the practical application of RT2text, a previously published rule-based method that converts summarized trajectories into human-readable descriptions. By using a public dataset of incidents from the Curitiba Municipal Guard, we showcase how RT2text can be employed to analyze the semantic nature of events over time. The method leverages regular expressions and templates to generate coherent narratives that facilitate data interpretation and insight generation, thus validating its effectiveness as a tool to support analysis and decision-making.*

**Resumo.** *A sumarização de trajetórias é um campo essencial para a análise de mobilidade, permitindo a extração de padrões de grandes volumes de dados. Este artigo demonstra a aplicação prática do RT2text, um método baseado em regras previamente publicado, que converte trajetórias sumarizadas em descrições textuais legíveis por humanos. Utilizando um conjunto de dados públicos de ocorrências da Guarda Municipal de Curitiba, mostramos como o RT2text pode ser empregado para analisar a natureza semântica das ocorrências ao longo do tempo. O método utiliza expressões regulares e templates para gerar narrativas coerentes que facilitam a interpretação dos dados e a geração de insights, validando sua eficácia como ferramenta de apoio à análise e tomada de decisão.*

## 1. Introdução

A análise de dados de trajetória é fundamental em domínios como planejamento urbano e segurança pública. Contudo, a complexidade desses dados exige técnicas de sumarização para extrair padrões relevantes. Embora eficazes, as saídas dessas técnicas frequentemente permanecem em formatos estruturados, de difícil interpretação para não especialistas.

Para lidar com esse desafio, o método *RT2text* (*Representative Trajectory to text*) foi proposto em [Menegol et al. 2025] como uma solução para converter trajetórias sumarizadas em descrições textuais. O *RT2text* utiliza uma abordagem baseada em regras e templates para gerar resumos concisos e legíveis, facilitando a compreensão de tendências de mobilidade. Este trabalho foca em demonstrar o potencial e a aplicabilidade do *RT2text* em um cenário real, utilizando um dataset público de ocorrências de Curitiba para gerar insights a partir da análise da variável semântica `NATUREZA1_DESCRICAO`.

## 2. Conceitos Básicos

A crescente prevalência de tecnologias de rastreamento levou a um aumento nos dados de trajetória, que capturam os padrões de movimento de objetos ao longo do tempo e

do espaço. A IoT e as mídias sociais expandiram isso para trajetórias de múltiplos aspectos (MATs), integrando aspectos espaciais, temporais e vários aspectos semânticos heterogêneos [Mello et al. 2019], como o modo de transporte ou o propósito da viagem.

Dado o enorme volume de dados gerados por GPS, a sumarização de trajetórias tornou-se essencial para a redução de dados, reconhecimento de padrões e interpretação humana [Machado et al. 2024]. Trajetórias Representativas fornecem uma abstração compacta dos padrões de mobilidade, melhorando o armazenamento, a visualização e a eficiência analítica [Machado et al. 2025]. No entanto, muitas vezes permanecem desafiadoras de interpretar, exigindo uma transformação adicional em resumos legíveis por humanos.

Para resolver isso, a Geração de Linguagem Natural (GLN), um subcampo do Processamento de Linguagem Natural (PLN), permite a geração automática de descrições textuais coerentes a partir de dados estruturados [Gatt and Krahmer 2018]. *RT2text* utiliza GLN baseadas em templates para transformar dados estruturados em narrativas legíveis, preenchendo estruturas predefinidas com os dados extraídos da trajetória.

### 3. O Método *RT2text*

O *RT2text*, introduzido em [Menegol et al. 2025], é um método implementado em Python que visa traduzir trajetórias representativas em linguagem natural.

A metodologia do *RT2text* é dividida em dois componentes principais:

#### 3.1. Pré-processamento

Nesta etapa, o arquivo de entrada contendo a trajetória sumarizada é carregado e analisado. Utilizando expressões regulares (regex), o método extrai e estrutura as informações relevantes, focando principalmente na descrição dos eventos de movimento. As características-chave, como as transições entre os diferentes valores do aspecto semântico (neste caso, a natureza da ocorrência), são identificadas. O sistema também classifica os eventos por períodos do dia (manhã, tarde, noite, madrugada) para organizar a narrativa final de forma cronológica.

#### 3.2. Descritor Textual

Com os dados estruturados, o *RT2text* gera o resumo em linguagem natural. Utilizando o motor de templates Jinja2, ele converte os padrões de movimento em sentenças dinâmicas. O método analisa as transições do atributo semântico para construir uma narrativa sequencial e fluida. Regras de formatação são aplicadas para eliminar redundâncias e melhorar a consistência do texto, resultando em uma descrição coesa e de fácil compreensão.

### 4. Estudo de Caso: Análise de Ocorrências em Curitiba

Para validar a aplicação do *RT2text*, foi conduzido um estudo de caso com o dataset de ocorrências da Guarda Municipal de Curitiba (SiGesGuarda)<sup>1</sup>. O conjunto de dados, fornecido pela Secretaria Municipal de Defesa Social e Trânsito (SMDT), compreende o período de 24 de novembro de 2022 a 25 de janeiro de 2023. Originalmente com 35 variáveis e 5.133 registros.

---

<sup>1</sup>Disponível em: <https://dadosabertos.curitiba.pr.gov.br/dataset/si-gesguarda>. Acesso em maio de 2025

## 4.1. Metodologia do Estudo de Caso

O estudo seguiu três etapas principais:

**1. Preparação dos Dados:** O dataset passou por pré-processamento para obter as coordenadas de latitude e longitude, com base nos endereços do registro de cada ocorrência. Nos casos em que a geocodificação do endereço falhou, foram usadas as coordenadas do respectivo Bairro de Atendimento.

**2. Sumarização da Trajetória:** As trajetórias diárias foram sumarizadas com o método MAT-SGT [Machado et al. 2025], gerando uma trajetória representativa para cada dia de ocorrências no dataset.

**3. Geração da Descrição Textual:** A trajetória representativa do dia 15/12/2022 foi usada como exemplo e submetida ao *RT2text* para análise da variável semântica `NATUREZA1_DESCRICAO` e conversão em texto.

## 4.2. Resultados e Discussão

A análise no *RT2text* focou nas transições entre os tipos de ocorrência com pelo menos 40% de representatividade em um determinado momento, produzindo a seguinte descrição:

Durante a madrugada, as ocorrências estão geralmente associadas a {TRÂNSITO (100%)}.

De 15/12/2022 00:07:46 a 15/12/2022 00:13:40, o foco mudou de {INVASÃO (50%), QUEIMA A CÉU ABERTO (50%)} para {PATRULHA MARIA DA PENHA (40%), APOIO (40%)}.

De 15/12/2022 00:08:23 a 15/12/2022 00:08:52, houve um deslocamento para ocorrências de {TRÂNSITO (50%), SUBSTÂNCIA ILÍCITA (50%)}.

De 15/12/2022 00:09:14 a 15/12/2022 00:09:50, a atenção se voltou de {TRÂNSITO (50%), SUBSTÂNCIA ILÍCITA (50%)} para {PATRULHA MARIA DA PENHA (67%)}.

De 15/12/2022 00:09:20 a 15/12/2022 00:09:50, a transição foi de {PATRULHA MARIA DA PENHA (67%)} para {INVASÃO (50%), FUNDADA SUSPEITA (ABORDAGEM) (50%)}.

De 15/12/2022 00:10:10 a 15/12/2022 00:10:48, o foco foi totalmente em {PATRULHA MARIA DA PENHA (100%)}.

\\ O restante da saída foi omitido por questões de espaço.

O resultado evidencia o potencial do *RT2text* em transformar dados brutos em uma narrativa interpretável. A descrição gerada permite que um analista, por exemplo, identifique rapidamente tendências como a concentração de ocorrências da Patrulha Maria da Penha no início da madrugada, o que pode direcionar ações específicas.

No entanto, a abordagem revela limitações importantes. O texto gerado, embora informativo, pode apresentar uma estrutura rígida. Mais criticamente, a análise sobre uma única dimensão semântica, a `NATUREZA1_DESCRICAO`, pode ocultar contextos essenciais. A sobreposição de eventos em intervalos de tempo quase idênticos ("De 15/12/2022 00:09:14 a 15/12/2022 00:09:50" e "De 15/12/2022 00:09:20 a 15/12/2022 00:09:50") é um exemplo disso: como as ocorrências aconteceram em locais diferentes, a distinção espacial se perde, o que reforça a necessidade de incorporar análises multivariadas para gerar descrições mais ricas e evitar ambiguidades.

## 5. Conclusão

Este artigo demonstrou a utilidade e o potencial do *RT2text* através de um estudo de caso prático com dados públicos. Ao converter uma trajetória representativa de ocorrências em uma descrição textual, a ferramenta provou sua eficácia em facilitar a compreensão de padrões complexos sem a necessidade de visualizações ou conhecimento técnico aprofundado. O estudo de caso com dados de Curitiba ilustra como o *RT2text* pode transformar resumos de dados em insights acionáveis.

Como trabalhos futuros, pretende-se aprimorar o *RT2text* para que possa processar múltiplos atributos semânticos simultaneamente. Planeja-se também automatizar os ajustes do texto para que ele se adapte dinamicamente a diferentes conjuntos de dados. Outras melhorias incluem a limpeza da saída para focar apenas em informações relevantes e a associação com outras variáveis, como a localização, para enriquecer a análise. Tais avanços visam consolidar o *RT2text* como uma ferramenta robusta de apoio à análise de dados e à tomada de decisões.

## Agradecimentos

Este trabalho foi apoiado pelo IFSUL (EDITAL PROPESP – N° 06/2024).

## References

- Gatt, A. and Krahmer, E. (2018). Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*, 61:65–170.
- Machado, V. L. et al. (2024). A survey on the computation of representative trajectories. *GeoInformatica*, pages 1–26.
- Machado, V. L. et al. (2025). Towards data summarization of multi-aspect trajectories based on spatio-temporal segmentation. *Journal of Information and Data Management*, 16(1):38–51.
- Mello, R. d. S. et al. (2019). MASTER: A multiple aspect view on trajectories. *Trans. GIS*, 23(4):805–822.
- Menegol, L. d. S., Machado, V. L., Bernardi, E. F. F., and Portela, T. T. (2025). Rt2text: Making trajectory summarization more accessible through text generation. In *Escola Regional de Banco de Dados (ERBD)*, pages 153–156. SBC.