# Review 70: [Short] Typical Decoding for Natural Language Generation

**Paper: https://arxiv.org/abs/2202.00666v6**

מאמר מסקרן המציע שיטת דגימה חדשה ממודלי שפה מאומנים. כמו שאתם בטח יודעים מקובל לדגום מילים (טוקנים) עם הסתברויות הגבוהות ביותר (בהינתן המילים הקודמות). ככה עובדים למשל nucleus ו- top-k. המחברים טוענים שפרדיגמת דגימה זו מובילה לטקסטים ״משמעמים מדי״ ויותר מדי דומים לדאטהסטים שהמודל אומן עליהם.

המאמר מציע פרדיגמת דגימה אחרת מבוססת על מילה טיפוסית (מבחינת האנטרופיה של המילה בהינתן הטקסט לפני) ולא על מילים הסבירות ביותר.בגדול מחשבים את האנטרופיה של מילה על סמך ההתפלגות שלה (softmax). לאחר מכן מחשבים את המידע של כל מילה (log2 של ההסתברות שלה). בסוף דוגמים מילים כאשר ההסתברות של כל מילה פרופורציונלי למרחק ההפוך של המידע שלה מהאנטרופיה של התפלגות המילה). כלומר ככל שהמידע של מילה קרובה יותר לאנטרופיה היא תדגם בהסתברות גבוהה יותר.

כך אנו דוגמים מילה טיפוסית (אנטרופיה היא למעשה המידע הממוצע של המילה) ולא המילה הסבירה ביותר.

יש במאמר דיון מעניין על אספקטים לינגוויסטיים של הגישה המוצעת.

סרטון של יאניק: https://www.youtube.com/watch?v=\_EDr3ryrT\_Y&t=2484s