## CÁC PHƯƠNG PHÁP ĐÁNH GIÁ CHẤT LƯỢNG BẢN DỊCH

Trên thế giới hiện nay, trong lĩnh vực dịch máy nói chung là có khá nhiều độ đo khác nhau để đánh giá chất lượng dịch máy. Chúng ta tạm chia các độ đo đó thành 2 nhóm dựa vào đặc điểm riêng của từng tiêu chuẩn: nhóm các tiêu chuẩn đo độ chính xác, nhóm các tiêu chuẩn đo tỷ lệ lỗi. Nhóm 1 có các tiêu chuẩn như: NIST (National Institute of Standards and Technology), BLEU (BiLingual Evaluation Understudy)… Nhóm 2 có các tiêu chuẩn như: WER (Word Error Rate), PER (Position-independent word Error Rate), TER (Translation Error Rate) ,… Trong phần này, tôi sẽ trình bày 3 phương pháp phổ biến và có hiệu quả được thế giới đánh giá là gần giống với chất lượng con người đánh giá: NIST, BLEU, TER. Trong đó, BLEU và NIST đại diện cho nhóm thứ 1, TER đại diện cho nhóm thứ 2.

### BLEU - BiLingual Evaluation Understudy

BLEU là một phương pháp dùng để đánh giá chất lượng bản dịch được đề xuất bới IBM tại hội nghị ACL ở Philadelphie vào tháng 7-2001. Ý tưởng chính của phương pháp là so sánh kết quả bản dịch tự động bằng máy với một bản dịch chuẩn dùng làm bản đối chiếu. Việc so sánh được thực hiện thông qua việc thống kê sự trùng khớp của các từ trong hai bản dịch có tính đến thứ tự của chúng trong câu (phương pháp n-grams theo từ). Phương pháp này dựa trên hệ số tương quan giữa bản dịch máy và bản dịch chính xác được thực hiện bởi con người để đánh giá chất lượng của một hệ thống dịch.

### NIST - National Institute of Standards and Technology

Phương pháp NIST là sự phát triển trên phương pháp BLEU nhưng có một khác biệt về quan điểm đánh giá là việc chọn lựa n-grams và thông tin trên mỗi n-gram sẽ được sử dụng để phục vụ việc đánh giá.

### TER - Translation Error Rate

Đây là phương pháp mới, vừa được công bố tại hội nghị uy tín AMTA 2006. Đây là một độ đo dùng cho dịch máy trong đó nó sẽ tiến hành đếm số lượng các hiệu chỉnh mà con người sẽ phải thực hiện. Vì vậy, phương pháp này rất gần với cách đánh giá của con người. Điểm số TER càng thấp thì hệ dịch càng tốt.