מבוא

כלים לעומק

שלב טרום ההמרה

ביצוע ההמרה של המילים לוקטורים יכול להתבצע באופן ישיר על הטקסט עליו אנו עובדים ובכך להשיג תוצאת התאמה בסיסית, ששומרת את המילים כפי שהם. הצעה נוספת הועלתה, לבצע מניפולציות על מילות הטקסט המקוריות בצורות שונות בשביל לגרום לוקטורים בפלט לשקף משמעות יותר כוללת של כל מילה. כל שיטה בשלב זה צריכה קודם כל לעבור את מבחן המשמעות של הטקסט בפני עצמו, עוד לפני ההמרה לוקטור ובחינת משמעותו. לכן בשלב זה נעבור על השיטות, ההגיון מאחורי הניסיון ואם נמצא לנכון נפסול בהתאם:

1. שמירת המילים כמות שהן- בעשיית מניפולציות ועל ידי כך שינוי המילים אנו בעצם משנים את הקונטקסט המקורי באופן מסוים ולכן שיטה זו, המשאירה את המילים במשמעות המדויקת בה אנו משתמשים, תיחשב עדיפה, כמובן בהנחה שתתן תוצאות זהות ומעלה מהשיטות האחרות. ואת זה נבדוק בהמשך.
2. הורדת רישות וסיפות ממילות הטקסט- על ידי סקריפט פשוט, עוברים על כל מילות הטקסט ובכל פעם שמזהים מילה בעלת רישא ו/או סיפא מורידים אותם ובכך יוצרים מילה שהמשמעות שלה נשמרת אך בקונטקסט שונה. למשל המילה כשאכלתי משמשת את הפועל אכל וזו משמעותה אלא שהרישא והסיפא הוסיפו קונטקסט של זמן ומדבר. אם נעבור על כל המילים ונוריד להם את הקונטקסט הנלוונ נוכל לקבל מילה שהמשמעות העיקרית שלה נשמרת והיא יכולה להתאים למילה דומה לה במקום אחר של הטקסט. במבחן התוצאה הטקסט שנוצר לאחר המעבר על רישות וסיפות כלל מילים חסרות משמעות כמו למשל מתוך המילה שלשום נוצרה המילה לשום שלה כמובן אין משמעות כלל. לכן שיטה זו נפלה במבחן טרום ההמרה.
3. הפרדת רישות וסיפות ממילות הטקסט- שיטה זו הינה שיטה משופרת לשיטה הקודמת. מצד אחד אנחנו כן מפרידים את הקונטקסט הנלווה למשמעות העיקרית של הטקסט, אך מצד שני לא מאבדים אותו לגמרי.
4. תייגן...
5. הפרדה לאותיות...

שלב ההמרה

תוצאות

מסקנות

סוף דבר ביבילוגרפיה