

תרגימא

גיליון מיוחד לרגל

כינוס אגודת המתרגמים בישראל

8-10 בפברואר 2010

FEB
2010



תרגימא

גיליון מיוחד לרגל כינוס אגודת המתרגמים בישראל 8-10 בפברואר 2010

אגודת המתרגמים רוצה להודות למחברים ששלחו את מאמריהם לגיליון מיוחד זה, לשרה ירקוני על מאמציה בשיווק ובהשגת החומרים, ליפית אלשייך ולחגית רוזנס על עזרתן, וכן לאינגה מיכאלי ולאסתר הלק על העריכה.

תוכן

ישראלית שפה יפה

פרופ' אורה (רודריג) שורצולד · 3

האם הדשא של השכנים אכן ירוק יותר? אינגה מיכאלי · 5

זיכרון תרגום – סקירה יוסי רוזנמן · 7

ישראלית שפה יפה

סקירה מאת: פרופ' אורה (רודריג) שורצולד, אוניברסיטת בר-אילן

מטרתו טל הספר **ישראלית שפה יפה** היא ״זלבדוק אם השפה שהישראלים מדברים היא אכן שפה שמית או שפה שהיא אירו-אסיאתית – גם שמית וגם הודו-אירופית. השקפת צוקרמן, המובאת בפרק ההקדמה היא שלידת העברית הישראלית היא ״הכלאה (היברידיזציה): ישראלית היא תוצאה של הצטלבות בין העברית (השמית), היידיש (האירופית) ושפות אחרות״ (עמי 12). עוד הוא טוען שהעברית והיידיש הן התורמות העיקריות לשפה הישראלית, אם כי הייתה תרומת מה גם ללשונות אחרות. שלושה חלקים לספר: החלק הראשון – ראשיתה של השפה הישראלית: עובדות ומיתוסים, ובו שלושה פרקים: ישראלית לעומת עברית; מקורות הישראלית; מיתוסים נפוצים לגבי העברית; החלק השני – ההשפעה הדקדוקית של היידיש ושפות אירופיות אחרות על הישראלית, ובו ארבעה פרקים: הגייה; מורפולוגיה; תחביר; אוצר מילים; החלק השלישי – ההשפעה המוסווית של היידיש ושפות אירופיות אחרות על אוצר המילים של הישראלית, ובו ארבעה פרקים: תרגום בבואה; תרגום שומר משמעות וצליל (תשמו״ץ); תשמו״ץ מילים מיידיש; תשמו״ץ מילים בינלאומיות. חותמים את הספר שני נספחים: המילונות הישראלית והמילונות היידית ומכתבו של גרשם שלום לפרנץ רוזנצווייג. ההערות הרבות מרוכזות אחרי הנספחים וביבליוגרפיה ענפה מובאת בסוף הספר.

לשון הספר קולחת אף שהוא מתורגם מן האנגלית, והטיעונים התיאורטיים משלבים דוגמות יפות. תוספות אנקדוטליות ובדיחות רבות, חלקן לעניין והרבה מהן שלא ממין העניין, וכן שירים וציטוטים מספרים, זרועים בתוך הדיונים, ובכך מנעימים את הקריאה. המוטו למרבית הפרקים מעוטר בציטוטים עבריים מקוריים ובציטוטים מתורגמים מספרות העולם. תרשימים רבים מבהירים את הדוגמות ואת כיווני ההשפעה שצוקרמן דן בהם. טבלת השוואה מובאת בסוף הפרק האחרון ובה 111 דוגמות למילות תשמו״ץ שמקורן בינלאומי, אם כי רבות מהן אינן קיימות היום. הלשונות המשווות הן אנגלית, יידיש, רוסית, פולנית, גרמנית וצרפתית (עמי 186–196).

לא אתווכח על השם שבחר צוקרמן לשפה – ישראלית ולא עברית ישראלית. עצם הבחירה, הנובעת מכך שאיטלקי אחד אמר לו שהוא מדבר ישראלית כי מקורו בישראל (וקרוב לוודאי שלא ידע שיש עברית) הופכת אצלו לאידיאולוגיה. אדון כאן בדברים עקרוניים אחדים שבהם לוקה הספר – והם רבים:

א. עצם הטענה שהוא משמיע על הרכיבים המוכלאים בעברית אינה חדשה. קולות התרועה שנשמעו בארץ כשיצא הספר לאור העמידו את ספרו של צוקרמן כמחדש ממש ולא היא. חיים רוזן בלקט מאמריו שהתפרסמו בספר ״שיחות על לשון והיסטוריה״ (1954), ובמהדורות אחדות של ״עברית טובה״ (מ-1957 ואילך) כבר טען אותם דברים, וכך גם חיים בלנק בסדרות מאמריו משנות החמישים של המאה העשרים שיצאו לאור לאחר מותו בספר ״לשון בני אדם״ (1989). ספרים אלו אינם נזכרים כלל אצל צוקרמן. לעומת צוקרמן, רוזן ובלנק בבקיאותם הבלשנית הענפה ובראייה הבלשנית הרחבה של העברית לא תלו את רוב השינויים ביידיש. גם זאב בן-חיים במאמרו המכונן ״לשון עתיקה במציאות חדשה״ (לשונונו לעם תשי״ג) נפקד מן הספר ודיוניו.

◀ פרופ' אורה (רודריג) שורצולד מרצה במחלקה ללשון עברית וללשונות שמיות באוניברסיטת בר-אילן, משמשת כיום כראש המחלקה. חברה יועצת באקדמיה ללשון העברית. כתבה ספרים על העברית ועל הלאדינו ומאמרים מדעיים רבים.

עיצוב · גרפיקלי, יונה קדרון-שלו

ב. על התגבשות העברית נכתבו מחקרים רבים מאוד, אבל המחבר מתעלם מהם בשיטתיות, ואזכיר רק מן החשובים בהם. ספרים: מאיה אגמון-פרוכטמן ועמנואל אלון "החייתאת העברית" (תל אביב תשנ"ה), נתן האפרתי "מלשון יחידים ללשון אומה" (ירושלים תשס"ד), ומאמרים: יהושע בלאו "הרהורים על תחיית העברית" (לשונו תשס"ג), שלמה מורג "העברית החדשה בהתגבשותה: לשון באספקלריה של חברה" (קתדרה תש"ן), שלמה יזרעאל "לתהליכי התהוותה של העברית המדוברת בישראל" (תעודה תשס"ב) ועוד.

ג. גם על השפעת היידיש על העברית נכתב רבות. מאמרו החשוב של בנימין הרשב "מסה על תחיית הלשון העברית" (אלפיים תש"ן) התולה ביידיש השפעה רבה, נעדר מן הספר. גם אזכור ספרו של נצר "עברית בג'ינס" (תשס"ח) חסר. יחזקאל קוטשר בספרו האנגלי על ההיסטוריה של העברית עומד בסעיפים רבים על השפעת היידיש על העברית. לו עיין בהם צוקרמן ביתר כובד ראש, היה מוצא אסמכתות רבות יותר, ולא דוגמות אנקדוטיות ואזוטריות.

ד. השפה המדוברת היא חזות הכול אצל צוקרמן, וגישתו היא הגישה האנכרוניסטית של "הניחו ללשון" שאפילו רוזן ובלנק לא קיבלו אותה במלואה. אין ספק שהלשון המדוברת חשובה, אין ספק שהשינויים הרבים בלשון מתרחשים עקב הדיבור בה, אבל הלשון המדוברת היא רק צד אחד בלשון. אצל צוקרמן אין לשון לימודים ואין לשון תרבות. הוא מבחין בין "בלשנים" יוקרתיים חוקרי הלשון המדוברת, ובין "לשונאים" שהם טהרנים, רובם סרי טעם, שעיסוקם בזוטות לא חשובות ולא עניניות לחקר הלשון. חלק מאותם לשונאים שהוא מזלזל בהם הם חוקרי לטון חשובים שטיעוניהם התיאורטיים ראויים לעיון, למרות הסתייגותם מן ה"שגיאות" בלשון המדוברת. ותמהתני איך היה בונה צוקרמן את התכנית ללימודי הלשון בישראל.

ה. בלשון המדוברת רבות המילים השאלות מן היידיש – בצד מילים שאלות מלשונות רבות אחרות, כפי שהראו מחקריי, וקיימים צירופי לשון רבים הקיימים ביידיש ומתורגמים לעברית, אבל חלקם קיימים גם בלשונות אירופה האחרות, ואין זה בטוח שהיידיש היא מקורם. צוקרמן לא עמד על כך שבלשון הדיבור, היא הלשון הנחשבת נמוכה במשלבים, יש תיעוד רב ליידיש. ככל שעולים במשלבי הלשון – היידיש נמוגה ונעלמת. ההיבט הסוציולינגוויסטי החשוב המבחין בין המשלבים והשפעות הלשונות אינו נרמז בספר: מלשונות בעלות יוקרה בעיני הדוברים התקבלו מילים שאולות ותרגומי שאילה רבים הנפוצים בכל משלבי הלשון במיוחד בלשון הבינונית והגבוהה; מלשונות בעלות יוקרה נמוכה בעיני הדוברים השתלבו מילים וצירופים מתורגמים רבים בלשון הדיבור, לכן מקום היידיש והערבית כה גבוה בלשון המדוברת.

ו. שני עקרונות לתיאוריית ההכלאה של צוקרמן: עקרון המייסדים ועקרון החפיפה. עקרון המייסדים אומר ש"מאפייני השפה של האוכלוסייה המייסדת קובעים את טבע השפה הנוצרת. לפיכך יידיש היא תורמת עיקרית לישראלית מכיוון שהיא הייתה שפת-האם של רוב מייסדי השפה ("מח" העברית") בשלב הקריטי של תחילת הישראלית" (עמ' 48). בעמ' 23 הוא מצי"ן את שנת 1900 כשנת ההתחלה של ה"ישראלית". ידוע הוא שהשפעת היידיש נפוצה בעקבות גלי העלייה השנייה (1904-1914) – שרבים מחבריה עזבו את הארץ והיגרו לארצות אחרות. למרות הסטטיסטיקות המובאות, הדוברים הראשונים של העברית (מ-1882 ואילך) לא היו דוברי היידיש דווקא אלא הספרדים שחיו בערים הגדולות ויצרו קשר עם אחיהם בני הקהילות האחרות. לא היחידים בעלי האידיאלוגיה כמו אליעזר בן-יהודה הם שתרמו

לדיבור העברי ההתחלתי אלא דווקא "עמך". רק עם בוא בני העלייה השנייה התגברה ההשפעה היידית, בעיקר בדיבור. העיקרון השני הוא עקרון החפיפה: "אם מרכיב בישראלית קיים הן ביידיש והן בשפה אחרת, יש להניח – אלא אם כן הוכח אחרת – כי הוא יובא מן היידיש" (עמ' 50). בין השפות התורמות האחרות הוא מונה רק רוסית, פולנית או גרמנית. לשונות אירופה האחרות אינן נחשבות, ולשונות היהודים לא כל שכן. יש הרבה מאוד צירופי לשון הקיימים במאגר האירופי הכללי וגם ביידיש, ואין זה בטוח שהמקור הוא יידי. כך למשל, ביידיש קיימת המילה 'מענטש' בהוראת "בן-אדם" בימינו. עצם השימוש ב"הוא לא בן-אדם" אינו יידי אלא ספרדי-יהודי, כי בספרדית יהודית נאמר "ו'ו איס בן-אדם" בהוראת "הוא לא בן-אדם". האם נאמר שזו השפעה יידית? למה נבחר דווקא השם "בן אדם" ולא "אדם", "איש", "אנוש", "ברנש" ועוד? ויכולתי לתת עוד כהנה וכהנה דוגמות.

צוקרמן מעלה טענות שאחרים שוגים בהן, למשל: "אנשים רבים מאמינים שהשפה אינה מכילה אלא רכיבים בלבד (מילים למשל) ומתעלמים מקיומם של מבנים" (עמ' 54). צוקרמן עצמו חוטא באמונה זו. הוא עוסק במילים יותר מאשר במבנים. מה שהוא כולל במקרים רבים כמבנים אינם אלא צירופי לשון מאובנים. לו עסק במבנים הוא היה יודע שסדר הרכיבים בצירוף השמני הוא: שם עצם, שם תואר, כינוי רמז – בדיוק כמו בעברית הקלאסית, ובכלל לא כמו ביידיש. היידוע הנשנה בשם העצם, בשם התואר ובכינוי הרמז הוא כמו בעברית הקלאסית ולא כמו ביידיש, וכך גם ההתאם התחבירי. המספרים (גם אם יש שגיאות במין) נמנים בסדר אלפים, מאות, עשרות, יחידות – ולא בסדר היידי ההפוך. הצרכות הפעלים אינן שוות תמיד למצוי ביידיש. הסמיכות ממשיכה להתקיים בעברית הן בצירופים כבולים חדשים הן בצירופים הנוצרים באופן חופשי מדי יום, ואין הם מומרים תמיד במבני "א של ב". הפרדת הכינויים הדבוקים מן השם ומן הפועל אינה תולדה של היידיש דווקא (למרות עקרון החפיפה). היא קיימת גם בלשונות אירופה האחרות, וראשיתה כבר בתקופות הלשון הקלאסיות של העברית. ועוד תופעות תחביריות רבות מאוד.

בהשפעה הפונולוגית למשל הוא מצי"ן את מערכת העיצורים. אמנם היגיו הצד"י כ-ts התקבל כמקובל ביידיש (ובאיטלקית), אבל לא הגיית התי"ו הרפה (כסס"ך). העדר הלועזיים והנחציים אינו "חודי רק לדוברי היידיש. הוא חסר גם בהגייתם של דוברי הספרדית היהודית, האיטלקית, הצרפתית ועוד. ההגייה שהתקבלה בסופו של דבר הייתה מה שאני מתארת "המכנה המשותף הנמוך ביותר", משמע מה שקיים במרבית הקהילות שהחלו להשתמש בעברית כלשון דיבור. חמור מכך הוא התעלמותו משאר מרכיבי המערכת הפונולוגית – התנועות והטעם. התנועות אינן התנועות שביידיש, והטעם לא כל שכן. הטעם מתאים למה שהיה מקובל בקהילות הספרדים בהגייתם העברית. לו הייתה השפעה יידית בולטת היינו מטעימים את **אפשה**, **משפחה**, **הקדוש ברוך** הוא ועוד. אמנם יש הטעמת מלעיל המבחינה בין שמות פרטיים לשמות כלליים, כגון **שושנה** – **שושנה**, רחובות-רחובות, וכן בין מילים אחרות כגון **כמה** –**כמה**, **ראשון**, **שני**-**ראשון**, **שני** ועוד. על אחדים מהם הוא דן וראה בהם רק השפעה יידית (אם כי יש לכך גם גורמים אחרים כפי שהראיתי במקום אחר). אבל הוא מתעלם מכך שמדובר בקבוצה קטנה של מילים, גם אם היא כוללת כמה עשרות ואולי מגיעה לכדי מאות מילים, והן מילות חיבה, שמות פרטיים, מילות משחק וקשָׁרים תחביריים – בעיקר מילות תפקיד. רוב רובן של המילים בלשון, עשרות אלפיהן, מוטעמות כמקובל בהגייה

ה"ספרדית" הממשיכה את מסורת ההגייה המקראית. כל הדוגמות במהלך הדיונים עוסקות כאמור במילים בודדות המזכירות את טענתו שהשפעת היידיש עצומה (למשל טביעת שם עברי מסוים במין נקבה כי השם ביידיש או בלשונות אחרות שמהן שאבה היידיש הוא בנקבה), אבל אין לצוקרמן הסתכלות מערכתית על מבנה העברית הישראלית, והוא אינו נותן לעובדות המוצקות על העברית בת-זמננו, המוכחות במאות מחקרים בלשוניים מובהקים (שאינם מוזכרים או מצוטטים, כמובן), לבלבל אותו עם התיאוריה. כך למשל הוא דן בסיומות -ניק ו-לה הבאות מן היידיש המזרחית, אבל אין הוא מזכיר כמובן את הסיומות -יקו, -ָה, -יסט, -יקה, -רייה, -יזם, -לנד ועוד שהגיעו מן המאגר האירופי הכללי.

י. לא ברור לי מה עניינם של הסיפורים הפלינדרומים (עמ' 89-90) או האיקוניות והספרות ההומופונית הדו-לשונית (עמ' 150-153) לענייניו של הספר. דברים אלו של המחבר התפרסמו במקומות אחרים והוא רצה ככל הנראה ליידע את הקורא העברי ברוחב מחקריו, אבל אין הם רלוונטיים לספר. גם לא ברורה הנחיצות של נספח 2 לספר הזה.

ולבסוף, ציפיתי לתיאור סינכרוני של העברית הישראלית, ותחת זה קיבלתי תיאור דיאכרוני-אטימולוגי, רובו כולל פרטים בודדים ומפוזרים שאינם מצטרפים לכלל מערכת שלמה. טוב היה עושה המחבר לו למד היטב את הנושא שרצה לחקור לפני פרסום הדברים.

ט.

האם הדשא של השכנים אכן ירוק יותר?

איננה מיכאל

בחודשים האחרונים נדמה שהתרגום זוכה לעדנה. מתרגמים "מפורסמים" ואנשי אקדמיה בכירים מרואיינים בתקשורת, שמו של המתרגם כמעט כבר לא נשמט כבעבר מביקורות התרגום המתפרסמות בעיתונות הכתובה והמקוונת, ראשי האגודה הזמנו לאחרונה לכמה דיונים חשובים בנושאים "חמים" כמו מכרזי התרגום, התמלול והמתורגמות עבור משרדי הממשלה (וגם לדיון ב"חוק הצרפתי" להגנה על הספרות העברית, שעלה לאחרונה לכתורות), ואף מצאו אוזן קשבת וראש פתוח.

נכון, זו רק ההתחלה והדרך עדיין ארוכה, אבל רוחות של שינוי החלו לנשב. השאלה הנשאלת היא: האם זה משפיע על מעמדו ו/או על מצבו הכלכלי של המתרגם, היושב ספון בביתו ושוקד על מלאכתו? והתשובה היא לא. רבות דובר בעבר על מעמדו של המתרגם, ועוד ידובר בעתיד לכשיתפרסם מחקרן המרתק של פרופ' מרים שלזינגר ושל פרופי רקפת סלע-שפי על ההביטוס ועל תפיסת המקצוע של המתרגם. אבל אם רק נווהו את מצבם של המתרגמים בישראל לזה של עמיתיהם בעולם – גם בלי לערוך מחקר של שנים – נגלה שהדשא של השכנים ירוק הרבה יותר, ושמקצוע התרגום עדיין לא זוכה להכרה הראויה.

בדברים שלהלן אתייחס בעיקר לשתי הקבוצות שנתפסות באופן פרדוכסלי כחלשות ביותר בעולם התרגום – אף על פי שהן גם המסוקרות ביותר והכי פחות שקופות (מעצם הקרדיט הניתן למתרגם) – והכוונה היא למתרגמי הספרות והמתרגמים לכתוביות. על-פי מקורות יודעי דבר, כמחצית מהספרות היוצאת לאור בישראל מדי שנה היא ספרות מתורגמת. בהצעת החוק שהניח לאחרונה על שולחן הכנסת ח"כ ניצן הורוביץ הוא אף מכיר בחשיבותה של הספרות המתורגמת – ולא רק המקורית – לתרבות הישראלית ולעיצוב זהותה, ומביע חשש מפני מצב שבו "בעוד שנים ספורות... ההיצע הספרותי יוגבל לרבי-מכר ולתרגומים קלוקלים".

בה בעת, גם חלק ניכר מכלל התוכניות המשודרות בשלל ערוצי הטלוויזיה שלנו מתורגם משפות זרות, בעיקר מאנגלית. התרגום הוא חלק אינטגרלי מחיינו גם בתחומים אלה וגם באינספור תחומים אחרים: מרפואה, דרך בתי-המשפט ועד אתרי אינטרנט רב-לשוניים שכולנו גולשים בהם מדי יום. כולנו צורכים תרגום. אבל אם תפנו לאדם מן השורה הוא לא יבין מדוע עליו לשלם עבור תרגום, על אחת כמה וכמה יותר משכר עבודתו של קלדן ממוצע.

אין בכוונתי לנסות ולנתח מדוע יש לאחרונה התייחסות כה רבה לתרגום ולחשיבותו, מצד אחד, ומצד שני מדוע סובל המקצוע מבחינה היסטורית מיחסי ציבור גרועים וממעמד נחות (גם בקרב המתרגמים עצמם, שאינם רואים בעצמם אנשי מקצוע או אנשי קריירה במלוא מובן המילה). לא, אני רק רוצה להביא כמה עובדות מהשטח, ולשם השוואה נעזרתי בנתונים על מה שקורה בעולם.

אז הרשו לי לסבר את אוזניכם, ונתחיל עם התרגום הספרותי.

מה קורה באירופה?

- במדינות רבות באירופה – בהן צרפת, בלגיה, שבדיה וגרמניה – התעריף הממוצע לתרגום ספר (שכר עבודה ללא תמלוגים) משתווה לרף הגבוה שמציבה אגודת המתרגמים בישראל (דהיינו: 1440 ₪ לגיליון דפוס; ראו גם: מה קורה בארץ); במקרים רבים התעריף שנחשב "טוב במיוחד" גבוה בהרבה מהתעריף המומלץ של אגודת המתרגמים.
- בכמה ממדינות אירופה התשלום על עבודת התרגום הוא

מקדמה על חשבון התמלוגים (בדרך כלל עד 2%). במקרים רבים אין לזה משמעות כי הספר נמכר במספרים זניחים, אך אם במקרה תרגם המתרגם רב-מכר, הוא יתוגמל בהתאם.

- ברבות ממדינות אירופה זכויות התרגום נשארות אצל המתרגם, ויש גם "סוכנויות גבייה" שדואגות לגבות עבורם את הכסף מהתמלוגים.

מה קורה בארה"ב?

- מספר הכותרים המתורגמים שיוצא לאור נמוך בהרבה מהנהוג באירופה ובישראל.
- מבחינת היחס למתרגם ולעורך וההכרה בערכם, המצב דומה למה שקורה בארץ: התרגום איננו פרופסיה ואינו זוכה להכרה. מתרגמים רבים נאלצים לעבוד במשרות נוספות (בעיקר באקדמיה) כדי להתפרנס.
- מתרגם ספרות משתכר על-פי מילה: 0.16–0.07 דולר למילה (דהיינו, 0.64–0.28 ₪ למילה). בהשוואה, מתרגם ספרות בארץ משתכר 0.3–0.075 ₪ למילה.
- נושא התמלוגים וזכויות היצרים (בשיעור של 3%–1) זוכה למענה במקרים רבים, אך רק בעקבות מו"מ בין ההוצאה למתרגם (ולעתים מתוך הבנה עם הסופר שהתמלוגים למתרגם יורדים מתוך חלקו שלו בהכנסות).

ומה קורה בארץ?

- מתרגמי הספרות בארץ מקבלים תעריף המחושב על-פי גיליון דפוס (ג"ד), דהיינו 24,000 תווים (או 16 עמודים מתורגמים של 250 מילה). התעריפים הנהוגים בארץ נעים בין 300 ₪ ל-1000 ₪ או מעט יותר לגיליון דפוס, (הרבה) פחות מהתעריפון המומלץ של אגודת המתרגמים). רק מתרגמים יזועים במיוחד (או מתרגמים לשפות נדירות יותר) מקבלים יותר. רוב מתרגמי הספרות בארץ מקבלים כ-700–600 ₪ לג"ד.
- עורכי הספרות מקבלים כמחצית מתעריף המתרגמים, דהיינו בין 500–150 ₪ לג"ד.
- הוצאות רבות אינן משלמות על רווחים (מדובר על הפרש משמעותי של 20% לרעת המתרגם).
- אורכו של ספר ממוצע כ-20 ג"ד. נכפיל זאת ב-700 ₪ לג"ד ונקבל תשלום ברוטו של 14,000 ₪. רוב המתרגמים עובדים על ספר בסדר גודל כזה כשלושה חודשים.
- עבור עריכת אותו ספר יקבל העורך כ-6000 ₪ ברוטו לעבודה של כחודש ימים.
- תנאי התשלום הממוצעים נעים סביב שוטף + 30 או 60.
- זכויות התרגום אינן נשארות אצל המתרגם, והוא איננו מקבל תמלוגים (ב-99% מהמקרים).

- שמו של המתרגם מתפרסם בדרך-כלל בעמוד הקרדיטים הפנימי, ושמו של העורך מתפרסם רק לעתים רחוקות מאוד (אף על פי שיש לו אחריות שווה לתוצר הסופי - עורך יכול לשפר פלאים תרגום בינוני או טוב - או לתרום את חלקו גם לתרגום מצוין - ועורך גם יכול להרוס תרגום טוב או מצוין).
- לשם הדגמה: מתרגמת ידועה, עם תואר שני בתרגום וותק של 20 שנה, מקבלת כיום מהוצאה X 900 ₪ לג"ד עבור ספר עיון מקצועי לחלוטין (פסיכולוגיה). בהוצאה Y היא עדיין מקבלת רק 650 ₪ לג"ד עבור ספר עיון חצי-מקצועי. אגב, אותה מתרגמת דיווחה שהיא עובדת בממוצע 10–8 שעות ביום.
- ומה עם התרגום לכתוביות, שאלתם? ובכן, שם המצב גרוע עוד יותר:
- עבור תרגום של סרט באורך ממוצע (90 דקות) יקבל המתרגם כ-100 דולר, ובמרבית המקרים יקדיש לתרגום בין יום עבודה מלא (אם הסרט קל במיוחד) ליומיים עבודה.
- התשלום הוא אחיד ללא קשר לרמת הקושי, והפרסטר היחיד הוא אורך הסרט/הסדרה, בין אם מדובר בסדרת נעורים קלילה, בסיטקום עמוס במשחקי מילים, או בסרט על מטוסי העתיד או על אסטרופיזיקה לערוץ נשיונל ג'יאוגרפיק.
- תנאי התשלום המקובלים ברוב אולפני התרגום: משוטף + 60 ועד שוטף + 105.

ואם נחזור לדשא של השכנים, להלן טבלה המבהירה עד כמה עגום מצבם של מתרגמי הכתוביות בישראל בהשוואה לעמיתיהם ממדינות שונות, מהודו ועד שוודיה. החישוב מבוסס על 15 פרקים של 45 דקות בחודש (כלומר בהנחה שהמתרגם יתרגם 180 פרקים בשנה). יש כמובן מתרגמים שעושים יותר ומרוויחים יותר, אבל זה הממוצע (לפחות בארץ). אם ניקח לדוגמה את הודו, נגלה שהכנסתו השנתית של המתרגם ההודי שיתרגם 180 פרקים בשנה נמוכה כמעט ב-40% מזו של הישראלי, אבל כוח הקניה שלו גבוה פי שניים כשמביאים בחשבון גם את יוקר המחיה. אם נשווה בין המצב באיטליה (גם מתרגמי הכתוביות האיטלקים מתאוננים קשות על מצבם), נגלה שהכנסתם גבוהה פי ארבעה מהנהוג בארץ, ואף עולה על התל"ג לנפש - בניגוד למצב בישראל.

ולסיכום הדרך ארוכה היא ורבה... כי נדמה שגם אנחנו המתרגמים וגם רוב לקוחותינו (מהוצאות לאור, דרך אולפני התרגום וכלה בקוראי הספרים ובצרכני הטלוויזיה) עדיין לא ממש הפנמנו: התרגום הוא מקצוע - חשוב ומאתגר - המצריך שנים של לימוד ושל התמקצעות. ועל איכות צריך לשלם.

המאמר הופיע לראשונה בגיליון ספטמבר 2009 של מגזין להב,בנושא "עולם גדול, עסקים קטנים"

	לסרט של 45 ד' (₪)	הכנסה לשנה (₪)	הכנסה (\$)	תל"ג לנפש (2008)	שווי כוח קניה (2008)	USD	באחוזים
הודו	140.00	25,200	6,300	1,070	589%	2,960	213%
סלובניה	1,800.00	324,000	81,000	24,010	337%	26,910	301%
צרפת	2,364.30	425,574	106,394	42,250	252%	34,400	309%
שוודיה	2,556.00	460,080	115,020	50,940	226%	38,180	301%
אלבניה	160.00	28,800	7,200	3,840	188%	7,950	91%
בריטניה	1,421.25	255,825	63,956	45,390	141%	36,130	177%
איטליה	872.54	157,058	39,264	35,240	111%	30,250	130%
ארה"ב	1,147.50	206,550	51,638	47,580	109%	46,970	110%
קנדה	630.00	113,400	28,350	41,730	68%	36,220	78%
ישראל	244.75	44,055	11,014	24,700	45%	27,450	40%

זיכרון תרגום - סקירה

מאת: יוסי רוזמן

אילו רק...

מתרגמים רבים מכירים את המצב שבו הם נתקלים בטקסט זהה או דומה לטקסט שתרגמו בעבר, או אפילו לטקסט שתרגמו רק כמה פסקאות קודם לכן. "יצא לי ניסוח כל כך מוצלח, איך לעזאזל זה היה?!" חושב המתרגם בייאוש ומנסה לדלות ממוחו היגע את אותו ניסוח חמקמק. או כמו שמתרגמים טכניים ומשפטיים רבים מכירים, טקסט שחזר על עצמו שוב ושוב. "אילו רק הייתה דרך להעתיק את התרגום אוטומטית".

מהצד השני של המתרס נמצאות חברות שנאלצות להשקיע כסף רב בתרגום מסמכים, מוצרים ו/או התייעוד הנלווה למוצרים. לעתים קרובות זהו רק עדכון של מסמך, או גרסה חדשה או משודרגת של מוצר קיים או דגם אחר, וחלק גדול מהטקסט זהה או דומה מאוד. "אילו רק יכולנו לשלם רק על הטקסט נטו, בלי כל החזרות", חושב לעצמו בעל החברה.

אז תארו לעצמכם שהיינו יוצרים טבלה, בעמודה אחת היינו מעתיקים את טקסט המקור ובעמודה השנייה את התרגום של כל טקסט שאנחנו מתרגמים. כל פעם שהיינו נתקלים בטקסט שנראה לנו זהה או דומה למשהו שתרגמנו בעבר, כל מה שהיה עלינו לעשות זה לחפש אותו בטבלה ולהעתיק אותו.

טוב, למלא טבלה, ואחר כך לעשות חיפוש ידני ולהעתיק את הטקסט זה די מייגע, לא בדיקו יעיל וגם לא ממש חוסך זמן, אבל אילו רק הייתה אפשרות לעשות את כל זה בצורה אוטומטית...

המחשב חש לעזרתנו

זהו בדיקו זיכרון תרגום (Translation Memory או TM בקיצור) - תוכנת מסד נתונים שאוגרת טקסט **שתורגם על-ידי מתרגם אנושי**, וכוללת אפשרויות חיפוש ואוטומציה מתוחכמות. כמובן שאין צורך להעתיק את המקור והתרגום לטבלה וגם לזכור אם תרגמנו משהו דומה או לא. התוכנה מזינה את המקור והתרגום למסד הנתונים וגם מבצעת חיפוש עבורנו, כל זאת אוטומטית. ולא זו בלבד, אפילו מוצאת משפטים דומים ומסמנת את ההבדלים בין הטקסט שאנחנו רוצים לתרגם לבין הטקסט במסד הנתונים. ממש ימות משיח.

כאן חשוב לעצור רגע ולהבחין בין תרגום מכונה (Machine Translation או בקיצור MT) ומילון מונחים או מונחון (Glossary) לבין זיכרון תרגום. שלושת אלה הם כלי תרגום בסיוע מחשב (Aided Translation או CAT), אבל קיימים הבדלים מהותיים ביניהם. **בתרגום מכונה**, התוכנה מנתחת את מבנה המשפט, ובאמצעות **מסד נתונים קיים** (שגם יכול לגדול עם הזמן) של מילים וצירופי מילים היא מתרגמת אוטומטית משפטים. בעצם, תוכנת התרגום האוטומטי אמורה לחקות את מה שאנחנו, בני האדם, עושים. דוגמה לתרגום מכונה הוא התרגום האוטומטי של גוגל. וזו דוגמה טובה מאוד, כי מי שניסה אי פעם להשתמש בפונקציה הזו יודע שבמקרה הטוב היא יכולה לתת רק מושג כללי מאוד על הכתוב בטקסט. אמנם קיימות תוכנות תרגום מכונה מתקדמות ומשוכללות מאוד, שאפילו מצליחות להגיע **לתוצאות די סבירות בתחומים מסוימים**, אבל הן עולות הון תועפות ולכן פותחו רק עבור השפות המדוברות ביותר. בכל מקרה, **הטקסט המתורגם חייב לעבור עריכה של מתרגם אנושי**.

מילון מונחים הוא רשימה של מונחים והתרגום שלהם, מילון מצומצם (יחסית) וייעודי למטרה מסוימת. זהו כלי חשוב מאוד

יוסי רוזמן הוא מתרגם טכני, כתב טכני ומומחה ללוקליזציה.

למתרגם, שמסייע לשמור על עקביות בשימוש במונחים, אבל מיועד למונחים בלבד לא למשפטים. רוב תוכנות זיכרון התרגום משלבות גם מילון מונחים לנוחות המתרגם.

אז אם כך, זיכרון תרגום זה...

לעומת תרגום המכונה ומילון המונחים, זיכרון תרגום אוגר משפטים שלמים ומחרוזות שלמות, ואת התרגום שתרגם מתרגם בשר ודם. כמובן, עד שלא הצטבר מאגר תרגום גדול מספיק, אין תועלת רבה בזיכרון התרגום. צריך לזכור גם שאיכות התרגום של זיכרון התרגום תלויה באיכות התרגום שהזן למסד הנתונים שלו.

יש גם להביא בחשבון שטקסט המקור צריך להיות בצורת קובץ, ובפורמט שניתן לעבוד איתו בתוכנת זיכרון התרגום - כלומר, שניתן יהיה לפתוח אותו במעבד תמלילים, עורך טקסט או בעורך של תוכנת זיכרון התרגום (בהתאם לתוכנה). רוב תוכנות זיכרון תרגום יזדעות לעבוד עם קבצים של מעבדי תמלילים (כמו Word), קובצי טקסט, HTML, XML וכדומה, ויש כאלה שאף יזדעות לפתוח קובצי DLL ,EXE ועוד.

לא רק שחזור ומחזור

זיכרון תרגום יכול גם לסייע לשמור על סגנון תרגום אחיד. לאנשים שונים ולחברות שונות יש העדפות סגנון שונות. אם לקוח חוזר אל מתרגם לאחר זמן מה, המתרגם יכול לדעת לפי מסד הנתונים איזה סגנון מועדף על הלקוח, ולתרגם בהתאם. נוסף על כך, מתרגם יכול להשתמש בתרגום שעשה עבור חברה אחת גם עבור חברה אחרת. אם לדוגמה תירגם עבור יצרנית מדפסות X, אין שום מניעה שישתמש באותו מסד נתונים לתרגום עבור יצרנית מדפסות Y. כמה שונים כבר יכולים להיות הטקסטים של אותם מוצרים?

זיכרון תרגום מאפשר גם עבודה של צוות מתרגמים על פרויקט תרגום גדול. המתרגמים מחליפים ביניהם מסדי נתונים, או עובדים על מסד נתונים אחד מקוון, כשבדרך כלל יש עורך שעובר על כל התרגומים, מתקן ומאחד סגנונות. ואם אפשר להחליף מסדי נתונים, למה שמתרגם ישתמש רק בתרגומים שלו ולא בתרגומים של מתרגמים אחרים? מתרגם יכול ליצור שיתופי פעולה עם מתרגמים אחרים ולהחליף איתם מסדי נתונים. בצורה שכזו הוא יכול לבנות מסד נתונים מכובד מאוד.

האם באמת הכול כל כך מושלם?

היו ודאי שישאלו את עצמם, מי ערב לכך שאיכות התרגומים שנקבל מהלקוח או ממתרגמים אחרים תהיה טובה? או אפילו תרגומים ישנים שאנחנו עשינו? הרי כולנו משתפרים עם הזמן.

מצד שני, נתקלתי כבר בחברות שצברו מסד נתונים עצום, מכל סוכנויות התרגום והמתרגמים שאיתם עבדו אי-פעם, ובנו למעשה מזבלה של תרגומים לא עקביים, ערבוביית סגנונות שונים, שגיאות תרגום, טעויות כתיב ושגיאות הקלדה. סיוט אמיתי.

חוץ מזה, אמרנו שלחברות שונות יכולים להיות סגנונות שונים, ולפעמים אפילו תרגומים שונים לאותו מונח. אם המתרגם ישתמש במסד נתונים אחד גדול לכלום, הבלגן יחגוג. אפילו באותה חברה ישנה אפשרות שישתמשו בסגנונות שונים למטרות שונות, למשל תרגום של ממשק תוכנה ותרגום טקסט שיווקי של אותה התוכנה.

ניהול ותחזוקה

התשובה לתסריט הבלהות לעיל היא **תחזוקה וניהול של מסד הנתונים**. כדי להפיק את המיטב ממסד הנתונים, חייבים להשתמש בו בצורה מתקדמת ולתחזק אותו בצורה שוטפת. תוכנות תרגום מתקדמות מצוידות במנגנונים מתוחכמים לסיווג,

as we needed them for our client projects and, slowly but surely, improved its capabilities. We looked at it as some kind of “secret sauce” that gave us the power to enhance quality and boost productivity.

Several years later we decided that we could also showcase its capabilities and use it as a marketing tool to promote our company’s brand. Under this initiative, we shipped ApSIC Xbench v2.1, a special version released as freeware that only supported the very popular Microsoft software glossaries. Since this was going to

be a public version, we were forced to review the user interface for greater clarity. Also the performance of the search feature underwent a thorough overhaul, because we were concerned about the impression the general public would have if it were not fast enough, given the size of Microsoft software glossaries.

From that initiative, we learned that making a version public significantly raised the bar on the quality of the tool, well beyond what is typically seen in internal tools. We also realized that greatly improving a tool our team had come to rely on had the

positive side effect of improving the overall execution of our own production processes.

A few months down the road, this positive experience led to the release of ApSIC Xbench v2.5. This release was not limited only to the Microsoft software glossaries. It was indeed the same version we used internally, with all bilingual formats that were supported at that time.

The release of ApSIC Xbench v2.7 set another important milestone when we added the now very popular QA functionality. It included QA checklists built on the search engine that allowed users to reach a new level of quality efficiency. In October 2009, we released the latest version, ApSIC Xbench v2.8. It added support for regular expressions, effectively bringing QA checklists to a new level thanks to the greater flexibility and accuracy of search expressions.

Our rapidly growing user base constantly pushes ApSIC Xbench beyond its limits, and the database of features to implement – both small and big – currently contains over 150 entries. The next milestone in ApSIC Xbench development that will be released soon is the full support of Unicode, a very important feature for multilingual projects.

Happy users tell us often that they think that ApSIC Xbench is worth money, but the current direction is for it to remain freeware, because it helps us reach a larger user base, and overall because we benefit from the great networking with translators and potential clients with a common interest in quality.

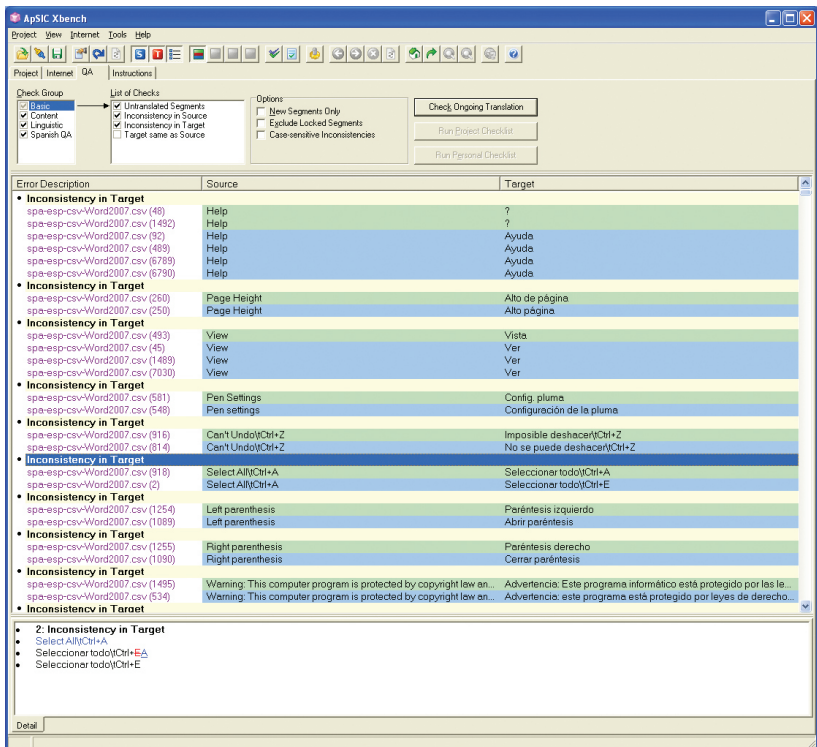
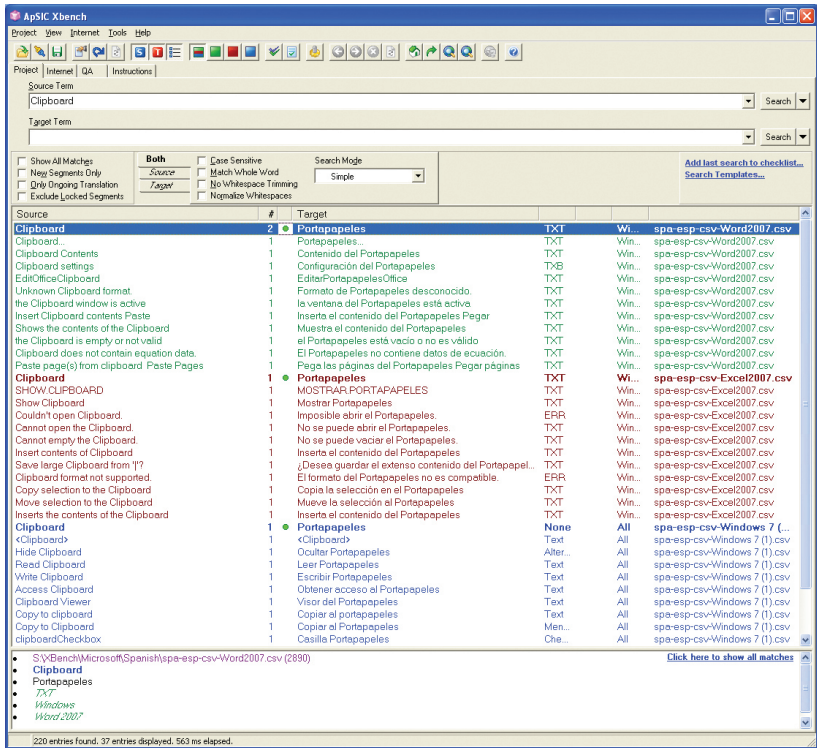
For more information about ApSIC or to download a free copy of ApSIC Xbench, please go to www.apsic.com.

למיון, לדירוג ולסינון של רשומות לפי מאפיינים או תכונות. בעזרת מנגנונים כאלה אפשר לדרג את הרלבנטיות של רשומה לפי שם חברה, נושא, הקשר ועוד, כך שרשומה עם מאפיין בעל ערך “X” תקבל עדיפות על פני רשומה זהה עם ערך “Y”.

בנוסף, כל מתרגם וחברה צריכים לקבוע לעצמם מדיניות לניהול מסדי הנתונים של התרגום. האם ליצור מסדי נתונים לפי נושאים, לקוחות, חברות, סגנונות? האם ליצור מסד נתונים אחד גדול או כמה קטנים? האם ליצור מסד נתונים נפרד לתרגום בעבודה ומסד נתונים נפרד לתרגום סופי?

סיכום

מאמר זה נוגע רק במספר היבטים של עבודה עם זיכרון תרגום, אולם אין ספק שבשימוש נכון כלי זה עומן בחובו יתרונות רבים, הן למתרגם והן לצרכן התרגום. המתרגם יכול לתרגם הרבה יותר מילים בזמן נתון ולהרוויח יותר, ואילו צרכן התרגום יכול לשלם פחות על טקסט שחוזר על עצמו או שתורגם בעבר ולחסוך כסף רב, שלא לדבר על שיפור איכות התרגום וקיצור זמני המסירה. לצד היתרונות, לזיכרון התרגום גם חסרונות, כמו עלות רכישה, עקומת למידה, סטנדרטים שונים ועוד, וכמובן חסדנות ובצע כסף של חברות שמשתמשות בו לניצול ולעושה של מתרגמים.



risk of error. The chances of catching mistranslations, wrong translations and other spelling and typos seem quickly balanced by the risk of inserting inappropriate usage of the references and guidelines, lacking knowledge of the subject matter or using different, conflicting interpretations. On the other hand, a conscientious translator going literally by the book by reading the whole document, compiling a glossary of terms and researching the subject prior to starting translating and with the help of a decent spellchecker can produce very satisfactory results. The end user pricing would subsequently decrease significantly as the overhead vanishes, and in these troubled times, this can possibly mean a budget available for more translations and therefore more translation work ahead.

Yes, but... if the translation can be improved by 3% thanks to additional layers of proofing, why then settle for less? How can we suggest that less than excellent is an option, since quality is paramount? Well, is it really? When the most sensitive services become simple commodities, in a demanding procurement environment where translation is considered non-strategic, it seems the "good enough" paradigm now prevails. The better often proves to be the enemy of the good. In the global village, generation X and Y customers value less quality craftsmanship than quick and dirty do-it-yourself, for it is cheap and yet serves its purpose. From fast food to disposable cameras, once life-or-death investments and decisions now give way to cheap and supposedly far less reliable alternatives, yet that are good enough. This observation extends to accounting principles since what until very recently was considered an asset and therefore subject to depreciation over 3 or 5 years can now appear in the expenses section of the Income Statement. A classic example is the generalized usage of the MP3 compression format. Diametrically opposed to Hi-Fi, this format has nevertheless literally conquered the world of digital music and is found and supported by any music player. Purists can condemn the quality, or rather the lack of it, but the facts speak for themselves, the goal is attained, the solution is good enough.

The same goes for translation, as for any other commodity. Beyond the red hot debate over the usage of crowdsourcing by for-profit organizations, the very fact that communities of users engage and take part in localization efforts for free confirm that strict quality assurance procedures are not a prerequisite for the development of a successful product. Users willingly forgive defects and bugs provided the product is available here and now and performs reasonably well. Less than perfect features can always be fixed later in flight.

If so, in order to market him/herself suitably, the successful service provider could in theory simply offer a good enough job. This mark of efficiency should not be misunderstood for sloppiness, for it is really the market that drives this call for "good enough". As a matter of fact, it is now quite common to observe a recurrent pattern in the offers. Along with the Super Premium Plus service level (what we used to call not so long ago standard, accepted, best practice) co-exist cheaper, lesser levels, with such names that are as misleading as they are poetic, Premium, Advantage, Business, Pro - the lower the level, the fewer quality mechanisms and validations involved - down to the proclaimed

usage of non-professional translators. Sure enough, professional associations invest heavily in educating clients in an effort to counter this trend. Such arguments as "would you let someone who just graduated from a six-hour first-aid course perform surgery on you?" resurface from time to time. As powerful as the image can be, can it really balance the commoditization of translation and the tsunami of crowdsourced free lunches?

Clearly, the answer is somewhere in the middle. Open-heart surgery will remain the prerogative of the distinguished practitioner and foot massage that of the guy next door. It all depends on the purpose. "Good enough" is the leading principle.

When Translation Memory systems started to pick up speed, the opponents among translation professionals were numerous and quite adamant. It was feared that the end of time had arrived, or at least the end of translation as we knew it. Yet, the battle was lost before it had been fought, and TM was imposed as a must, at least in technical translations. Quite anachronistically, I recently received an email from a translator requesting to offer his services and among the various qualities he claimed, he stated: "I never use any CAT tools; their use is against my philosophy of translation as an art rather than a semi-automated process." Well, it is true, TM is a flawed system. TM have countless limitations and drawbacks. They have little value per se, if any at all. Without context information, they are merely a collection of tentatively equivalent phrases whose validity can be questioned. Perfect matches derive their value from their applicability, which is intrinsically linked to context. Moreover, without strict control over the inputs, translation memories can quickly become a big bucket of useless garbage. Nevertheless, when used properly, TM can improve turnaround times and consistency, both vertical (over time) and horizontal (between parallel product lines). TM have not deprived translators work, on the contrary, they have allowed demand to be more demanding and volumes to increase.

In the quest for ever-faster turnaround, generalized simultaneous shipment and as volumes increase even more, crowdsourcing, whether or not on a voluntary basis, has already proven itself. From Harry Potter to Facebook, several translators working in parallel produce results much faster. Granted, but at what price? Well, it is definitely not as polished and error-free as with only a single translator, or at least a smaller team of translators. In many cases, bugs will be debugged along the way. Meanwhile the translation will serve its purpose, the overall result will be good enough and the users will overall be happy.

To the dismay of many hardcore purists, there are even more changes just around the corner. As technology evolves and computing power increases, one can only be baffled by the impressive results of Machine Translation. The various systems, whether rule-based, statistical or combined, are already here. They are operational and already integrated in major leading CAT tools. TM is already a declining mindset, MT is becoming white hot. Once translators get passed the incredulous smile and concern for the future of their craft, they cannot ignore that the revolution is in full throttle. Those who do not join may be fighters of a heroic cause, but they will be left behind. Adaptation is the key. MT, as TM before it, as computers before that and most likely typewriters in their time, is just a tool that can be used for the best as well as

for the worst. The demand is here for more and faster content. By combining the artificial intelligence of the machine with the knowledge of a seasoned expert translator, we may even be able to solve the impossible speed + price + quality conundrum.

Happy 2010.

Apsic Xbench, the Small Search Macro That Became a Localization Qa Product

ApSIC, the developer of ApSIC Xbench, is a Language Service Provider based in Barcelona, Spain, focusing primarily on translation and localization for the Spanish market. Since ApSIC's early days, we have had in-house software development skills, and because of that, we have frequently written our own tools whenever we spotted an important missing feature in commercial CAT software.

ApSIC Xbench is a powerful tool consisting of tens of thousands of lines of code to tackle two main functions of the localization process: Terminology Search and Quality Assurance. In ApSIC Xbench, the goal for Terminology Search is to provide terminology results very quickly and to do so in a useful and relevant format that reflects the expectations, background and needs of professional translators. The goal for Quality Assurance is to quickly provide lists of potential translation issues for areas that are not currently covered by most CAT tools, while trying to keep false alarms to a minimum.

However, ApSIC Xbench's features are not really the implementation of an initial master design, but they instead are the ongoing evolution a few very basic ideas and principles pushed by user feedback over the years. Actually, it is quite amazing to us that both functions originated in two small, quick and dirty Microsoft Word macros with roughly a couple dozen lines of code each. Back in 1999, when translating online help, we often struggled to obtain a fast and precise answer to a simple question: "Was that software item referred text already translated and, if so, what is the exact translation?" We came up with an arrangement of reference software strings in an MS Word document, color-coded to indicate priority – just like in the current ApSIC Xbench implementation – and we wrote an MS Word macro with just a few lines of code to search for the highlighted text in the bilingual reference document with software strings. That macro turned out to be so useful that we soon got used to the convenience of having relevant terminology at our fingertips.

However, when translations had to be done outside of Word – such as in SDLX, TagEditor or IBM Translation Manager – we felt left out in the cold. The reaction to that sudden need was ApSIC Xbench 1.0. It allowed us to search for terminology with a very clear view of priorities from any Windows application by simply pressing a certain sequence of keystrokes (Ctrl+Alt+Ins). Since the main ideas of the original design came from the concordance feature that is available in most commercial CAT tools, when we tried to describe what ApSIC Xbench was, we often defined it as a "concordance on steroids".

Initially, ApSIC Xbench was our own internal tool for terminology search and supported very few formats. In the equation Translation Time = Thinking Time + Typing Time, we realized that ApSIC Xbench allowed us to reduce Thinking Time dramatically. This perception of a major improvement led to a very high user buy-in, so it very soon became a standard tool for all of our translation projects. Over time, we added bilingual formats

As an Israeli who is a leader in the American Translators Association, it has been a joy for me to watch how the Israel Translators Association has evolved and taken its rightful place among the translation organizations in the international arena, bringing kavod both to Israel and to the translation and interpreting profession. The logical and most beneficial next step would be the creation of an entity to represent LSPs within the ITA.

Seven Free Tools to a Make a Translator's Day

By Naomi Elbinger

You can improve the speed and quality of your translation services with these seven smart tools, which are all free and easy to use:

- 1. Langover** – akou! We all know what akou means – it's the official email greeting of translators who forgot to switch their keyboard from Hebrew to English. If you have ever typed half a sentence only to look up at your screen and realize it's gibberish for this reason, then you need Langover. This is nifty tool gives your computer a hotkey that will translate your gibberish into the desired script. Langover supports all language pairs.
As a bonus, Langover also gives you a new shortcut – Ctrl+G – which allows you to instantly Google a word without switching to a browser. I feel sure that whoever invented this time-saving godsend must have been a translator.
<http://langover.com/>
- 2. Keyboard shortcuts** – Time Management gurus will tell you that the best way to start improving your work productivity and efficiency (and avoid Repetitive Strain injuries) is to learn keyboard shortcuts. Allow me to save you several hundred shekels + VAT by offering this advice for free!
In general, the more shortcuts you use the better, but there are several shortcuts that all translators should be using, such as switch languages (Alt+Shift), switch alignment (Ctrl+L, Ctrl+R), toggle caps (Shift+F3), and others.
There are lots of listings of keyboard shortcuts on the internet, so I'll just offer one below. Make your own list of shortcuts you think will be useful. Soon they will become second nature and you'll wonder how you managed without them.
http://en.wikipedia.org/wiki/Keyboard_shortcut
- 3. Desktop Search** – This is for sure the fastest way to find any document, email, or attachment that has ever entered your computer's memory. Using the toolbar, just start typing an unusual word that you know appears in the file, and Desktop Search will likely find the file for you before you even finish the word.
This tool works just as fast whether the file entered your computer's memory five minutes ago or five years ago, and even if it is stored on external hard-drives or in obscure, inaccessible nooks of the memory created when you transferred data from old computers.
<http://www.microsoft.com/windows/products/winfamily/desktopsearch/default.msp>
- 4. Translator's Abacus** – This is a free word-count program. It's supposed to do word counts in PDF, which would be great for giving estimates, but I and other translators have

found sometimes it unreliable in this capacity, especially on Hebrew PDFs.
However, Translator's Abacus is useful for doing word counts on multiple Word documents in a single click. That means that if you have a client for whom you perform ongoing translations, you can easily tally up the total word count in order to send them a bill at the end of the month.
Translator's Abacus is particularly time-saving if you have several files containing an unknown number of words. Rather than opening each one up, doing a word count and then adding that number into your formulized invoice (see below), use the Abacus to perform the word counts without opening the documents.

<http://www.globalrendering.com/download.html>

- 5. Kitzur** – This is an oldie but just so beloved that it deserves a mention! Israelis love to use acronyms and so will you, once you make friends with this dictionary of Hebrew Acronyms. Forget the time it will save you, just think of the brain-fatigue.
<http://kizur.co.il/home.php>
 - 6. Invoice templates** – If the end of the month means a struggle to adapt your clunky home-made invoice to send out to your clients, then it's time to get a template. Just add in your personal details, save the template and voila! Your personal invoice stationery is ready to use. It includes formulas, so that will also save you some time.
These invoices are easy to read and look professional and, in my experience, clients love them.
<http://office.microsoft.com/en-us/templates/CT101172551033.aspx>
 - 7. Zap – Zap** is a little program that works like the Find and Replace feature in Word, only it allows you to perform up to 10 commands in one Zap. This is certainly a huge time-saver when editing long translations for style and consistency. For example, you can replace all double spaces, misspelled names and transliterations that, on second thought, you decided to translate, in a single click.
The problem with Zap is that, unlike Word's Find and Replace, you can't use it to change formatting. (I wrote to the company about this and they conceded the problem but offered no solution.) Still it is certainly a time-saver in many cases.
Zap was developed by t4t – Productivity Tools for Translators. They have quite a few smart programs on offer, some for free and some for pay.
<http://www.itla.ch/t4t/tools.php?tool=zap>
- I hope that everyone finds something new and useful on this list. Remember: minutes that you save will add up to hours, which will add up to free money at the end of the month!
If you know of other tools, please write to me at naomielbinger@gmail.com

The 2010 Translator's Horoscope: Get Ready for More Change

With new paradigms entering the scene to meet new needs as well as answering old questions, the keystone of today's translation industry is being dislodged and risks swift replacement. Once reserved for an unlikely, science fiction future, MT may simply be the next big thing ahead. Will translators know how to adapt?

By Ilan Bloch

In a recent series of eye-opening statements, Renato Beninato, industry analyst, visionary, and to a large extent a mentor for many of us in the Language Service industry, pointed out a few myths on which the language industry has developed over the past couple of decades. One such myth is the widespread, transmitted belief that the more eyes involved, the better the quality. No self-respecting localization process dares depart from the holy trinity of Translation, Editing, Proofreading and its coined acronym TEP. However, linguistic and functional QA workflows alike are based on the assumption that after a while the tester just cannot physically see the errors and a new, fresh pair of eyes is needed.

This workflow of several passes is not only taught and professed by just about everyone everywhere, it is also at the heart of the profession's main concern, namely Quality. Regularly, in specialized press, dedicated blogs, and industry conferences around the world, quality is the most recurrent theme, from methods of measurement and descriptions of quality to processing in order to ensure that quality is indeed going to happen. In an effort to standardize processes and bring more objective comparison criteria in a lively industry, the powers that be devised and implemented such international standards as ISO 9001 or EN15038, which focus essentially, if not entirely, on quality.

Predictably, agreeing on the definition of quality is by and large not an easy challenge. Every so often, new metrics are proposed whose formulas include an ever more complex matrix of parameters to produce the score that will beyond any doubt objectively certify the quality level of the translation. Of course, subversive minds could question the true impact of the editing, proofreading, reviewing, revising, proof reading and other validating of the material, especially from a ROI perspective. If the gross cost of the translation is increased by 30% due to additional pairs of eyes going over the material, is the result going to be 30% better? Inversely, is it justified to increase the production costs by 30% to get only a 3% increase in the output quality? But these considerations involve the very measurements and definition of quality that are so difficult to agree on in the first place.

However, another approach proves even more fatal to the theory. Human intervention means by definition risk of error. As the French romantic poet Théodore Banville put it "ceux qui ne font rien ne se trompent jamais" (those who don't do anything never make mistakes). Each intervention - or perhaps interference - in a translation project by essence increases the odds of errors. The more passes, the more people involved, the greater the

decisions. This site also offers an aligned concordance of Hansard, the proceedings of Canada’s House of Commons, in French and English, and a Spanish-English collection of documents made available by the International Labour Organization (ILO)!

The holy grail of all mind-numbed and confused revisers is a software application that will isolate, at least, the most egregious errors. Such an application is actually being developed in Canada: <http://rali.iro.umontreal.ca/Traduction/TransCheck.en.html>. It can check for terminological consistency, false cognates, calques, illicit borrowings and other glaring translation errors between English and French. Such an application relieves the reviser of the mindless and mechanical aspects of editing and, of course, ensures that such errors do not go unrevised.

Identifying and correcting unidiomatic syntax is an especially daunting task. Again, the surest path to a successful self-revision or edit is a sound theoretical knowledge of your two language pairs. Thus, foreknowledge of likely errors and a checklist to refer to during your read through will bolster your faltering intuition. Hebrew, for instance, tends to use the passive more than English does. French tends to indulge in rhetorical questions and expresses points in the negative more often than English does. So the overuse of the passive voice, rhetorical questions, and points expressed in the negative should raise a red flag in an English translation, especially since these constructs impact differently on an English reader than they would on a Hebrew or French one, respectively.

While it is difficult to use Google to check correct syntax since you must select a particular wording (though you can search for phrases such as “on the one hand...but, on the other hand”), there are special databases of English texts that have been grammatically parsed. Oftentimes, however, these are difficult to use. They can be found by Googling “International Corpus of English”.

The holy grail of unidiomatic syntax revision may be another software application that can actually find errors you might not be aware: this is the standalone style checker, such as Style Writer: the Plain English Editor. Although these checkers were not designed to help translators with questions of unidiomatic syntax, they highlight problematic sentences and offer recommendations that may solve a translator’s problems.

Ultimately, even in the Internet age, the best way to identify and correct these types of errors in your own work may still be to have a standing arrangement with a colleague to read each other’s work or to consult with a colleague if you are unsure. Failing this, you can rewrite any sentence that seems suspicious, but if you are editing someone else’s work be sparing with the red pen because you may be bringing your own idiomatic, syntactical, or stylistic idiosyncrasies into play!

A New-Found Craze

By Gila Blits

I've played Scrabble for years at the Tel Aviv club. But I've now taken it one step further – I've started going to international tournaments. So far I've been to two, both of them in Malta. And the interesting thing, for me, is that I do better at these tournaments than I do in Tel Aviv. At the club, I'm in the second division (we're about 30 regular players in three divisions) and currently I'm winning 64% of my games (but my year on year average is probably nearer 50%). So at the moment I'm 12th in the club. Yet in Malta, where quite a few of the players have very high international ratings, I ended up 10th (last year) and 12th (this year), both times out of 60 players.

Scrabble words are often a matter of suspending disbelief. You can find the most peculiar words, especially those with two and three letters, but the compilers of the Scrabble dictionary claim that only words found in at least three well-known dictionaries are included. Lots of Hebrew words are acceptable, such as kiddush, kaddish, ovel and avel, eruv, erev, mazeltov (yes, it's one word for us!), shabbatot (but no shabbat), and since these are easy for Israelis to spot on a rack of seven tiles, it gives us a slight advantage. And of course, everyone always tries hard to make 7-letter words, or even 8- or 9-letter words by exploiting a letter or letters already on the board. These are the bingos, for which you earn a bonus of 50 points.

At the Malta tournament we play 20 games over three days. Players come from places as unexpected as Nigeria and Romania, as well as England, Ireland, the U.S., Australia, and of course, there's the Israeli contingent – this year we were six. We have to use clocks (like chess clocks), which are pre-set to 25 minutes per player. Exceeding your time results in a 10-point penalty per excess minute, which is deducted from your final score. Chatting is discouraged during a game, but as everywhere, there are those who find that difficult. The atmosphere "at the top" is fiercely competitive, but lower down the order it's fun, and it's great to meet new people.

Maltese is an interesting language, with roots in Arabic and lots of words adapted from Italian and English. It's extremely hard to read, mainly because it uses the letter X in ways unfamiliar to us. But since everyone speaks English, we haven't had to exercise our minds too much in that direction.

Next on my overseas agenda is a tournament in Coventry. I'm still a little hesitant about this one, because apart from the expense (and the fact that it's in January, which is no time to go to England), I worry that I won't have the stamina to get through it – it's 40 games over five days. Can I do that? I'm a great-grandmother, after all.

- Gila Blits has been a freelance Hebrew-English translator since 1986, and specializes in financial and legal texts.

Translators, Translation Companies and Professional Associations: Why Collaboration Benefits Everyone

By Rina Ne’eman

My experience with the Israel Translators Association is as somewhat of an observer. I am an Israeli, and have deep professional and personal ties to Israel. My company focuses primarily on Hebrew translation and related services. While we work extensively with translators who are members of the ITA, the fact that I am based in the United States means that my “ITA experience” is limited primarily to the annual conference and I miss out on the myriad of ongoing professional events that take throughout the year.

I have watched with great admiration and respect at the manner in which the ITA has evolved in recent years. From a conflict-ridden and marginal organization, the ITA has grown into a cohesive, well-run and greatly respected professional association that offers tremendous value to its membership. The ITA leadership has worked hard, and has clearly reaped the fruits of its labor – to the benefit of all translators and interpreters in Israel.

While I do not have close-up knowledge of the inner workings of the ITA, I am quite proficient in and familiar with international translation and localization organizations, including those that are geared primarily toward translation companies or language services providers (LSPs). In my capacity as head of the Translation Company Division of the American Translators Association (ATA-TCD), I have been asked to offer some insights as to why I believe that a similar entity would be beneficial to the ITA, to Israeli translation companies and to Israeli translators and interpreters.

Some might ask – why bother? Others might be concerned about what they view as opposing interests – translator versus translation company. The ATA model shows that it is in everyone’s interest for all of the major stakeholders in the translation and localization industry to be represented. It is my belief that this is best facilitated within the framework of the same entity. Other industry organizations in the international arena represent translation companies only – but by combining freelancers and translation companies within the framework of a single professional association, everyone’s interests are better served and optimal collaboration is far more readily achieved.

The ATA-TCD Model

The Translation Company Division is a relatively new part of the ATA. Although the ATA just celebrated its 50th anniversary – and is the largest translators’ organization in the world – the Translation Company Division recently marked its tenth year.

Ten years ago, the leadership of the ATA understood that

- Rina Ne’eman is a Hebrew translator & interpreter and managing director of Rina Ne’eman Hebrew Language Services. Rina works with prime ministers, major television networks and top companies and law firms worldwide. She heads the Translation Company Division of the American Translators Association.

the world was entering a new era, and that the translation and language services industry, led by LSPs, would be the most crucial component of global business and communication. The ATA-TCD was created as a framework that promotes the unique concerns and interests of LSPs and as an environment that facilitates successful collaboration among translation companies, as well as between translation companies and freelancers. The establishment of the ATA-TCD enabled us to reach out to each other, to communicate with each other, to share ideas, to learn and to network. The ATA as a whole is far richer and better off by its inclusion of translation companies, within its own, dedicated framework.

An LSP Division for the ITA?

An LSP division within the ITA potentially would be worthwhile for everybody – translation companies and freelancers alike. To be sure, the translation market is not a zero-sum game, in which freelancers fall victim to exploitative translation companies. That mindset is contrary to good business. Translation companies are the ones who link freelancers to work, manage the projects, serve the client, and make sure that the job is done right. For a quality-oriented translation company, translation is only one component of a job well-executed, preceding thorough editing and proofreading, and incorporating solid project management. Translation companies can provide freelancers with regular work, and eliminate much of the non-billable time spent by freelancers on activities such as marketing, client interface and collections, freeing them for the more lucrative task of actually translating.

An LSP division within the ITA would clearly benefit Israeli translation companies. Our experience in the ATA has been that translation companies need a framework in which they can come together and benefit from the advice of colleagues and experts whose knowledge can help to develop the translation and localization industry and thus to expand the market. There is no question that this applies to the situation in Israel, as well.

An independent LSP division could hold its own conference each year, geared specifically toward the interests and concerns of translation company owners and senior managers. In addition, special division sessions could take place within the framework of the annual ITA conference. This would help translation companies to develop skills related to project management, human resources, industry standards, technology decisions, machine translation and general industry trends, on a company and industry level. Moreover, this would create an opportunity for Israeli LSPs to forge strategic partnerships that are conducive to good business.

The “professionalization” of our industry is in everyone’s interest – translators and translation companies alike. To be certain, not doing so means continued poor public perception of translators and interpreters. A lack of understanding on the part of the public means diminished image, lower rates, and fewer opportunities for truly worthwhile work. The creation of an LSP division would open up new opportunities for all translation and localization professionals in Israel, by enhancing client education, leveraging public relations on behalf of the industry and creating frameworks for ongoing education on topics geared primarily to LSPs.

The **Corpus of Contemporary American English** (COCA, available at <http://www.americancorpus.org/>) is the largest freely available corpus of English. COCA contains more than 400 million words of text divided between spoken language, fiction, popular magazines, newspapers and academic texts. The corpus is updated annually. You can search for exact words or phrases, parts of speech, segments of words, and combinations of these. You can also search for surrounding words within a predetermined word span. Genre and time are two other features on which you can base your search. The website has a five-minute guided tour available to anyone who wants to learn more about the major features of COCA.

TIME Corpus (available at <http://corpus.byu.edu/time/>) is a 100-million-word corpus of texts written in American English from 1923 to the present, as found in TIME magazine. Aside from phraseology and grammar topics, you can search for how words/ parts of words (prefixes and suffixes, for instance) have changed meaning/frequency over time in authentic texts. If you are dealing with a translation of an old text, and want it to sound as if it belongs to the period, you can have the corpus generate a list of words that were used more in a certain period than another. For example, you can find nouns/verbs/adjectives whose usage increased/fell in a certain decade or period.

The **Ireland Component of the International Corpus of English** (ICE-Ireland) provides 300 transcribed spoken texts in 15 different discourse situations. It covers formal and informal discourse situations and includes almost 1,000 speakers. It also includes 200 written texts from 17 published and unpublished sources. It is perfect for linguistic research on present-day Irish English. As the website states, if you are a bona fide academic researcher and wish to obtain a copy of the ICE-Ireland CD-ROM and/or the ICE-Ireland Handbook for your private research purposes, copy the link <http://www.qub.ac.uk/sites/ICE-Ireland/HowtoObtainICE-Ireland/> on your browser and download and print out the form. Then sign and return to the address given on the form, together with the required fee.

Some Hebrew corpora:

The **Hebrew Corpus of Arutz7 Newswires** contains news and articles from Arutz 7 from 2001, which are updated daily. Researchers interested in downloading the content simply have to complete a registration form to receive the corpus in plain text format. A medical corpus, a corpus of spoken Israeli Hebrew and a corpus of economic texts are just some of the other corpora which can be downloaded from Mila Knowledge Center site at <http://www.mila.cs.technion.ac.il/english/resources/corpora/a7corpus/index.html>.

The **Hebrew Corpus** (available at http://www.cs.technion.ac.il/~saib/unambig_heb_corpus) is a project developed at Ben Gurion University. The corpus has been adapted to MorphTagger unambiguous file format.

If you would like to have further information about translation and corpora, here are some useful links:

Perrotti-Garcia, AJ. Customised corpora – a source of information for translators: ITI 2007 Conference oral presentation. Available at:

<http://www.iti-conference.org.uk/itic21/sessions/ITI%202007%20conference%20Perrotti-Garcia%20presenta.pdf>.

Izre'el, S. et cols. Designing CoSIH: The Corpus of Spoken Israeli Hebrew. International Journal of Corpus Linguistics, Vol 6, Number 2, 2002, pp. 171-197(27) Available at <http://www.ingentaconnect.com/content/jbp/ijcl/2002/00000006/00000002/art00001>.

This article is a brief and practical introduction to corpus studies for translators. If you have any questions or comments, please feel free to contact us at the email above.

Revising and Editing Translations: Must We Admit that “This Is the Situation and What Are the Chances It Will ever Change”?

by Meshulam Gotlieb

As translators, one of the most difficult jobs we face is revising our own translations or editing poor translations created by others. In particular, removing those last vestiges of the source language, primarily its idiom and syntax, that have invaded what should be a pristine target-language document, can be the most frustrating task. As one, veteran, Canadian translator, Brian Mossop, has noted, “When people are translating into their native language, they often write ungrammatical and especially unidiomatic sentences, under the influence of the source text... translators – even good ones – are prone to producing, under the influence of the source text, unidiomatic combinations such as ‘washed his teeth’”(Mossop 42, 136).¹ Indeed, in some cases, the translations given to us to edit are so poor that the best solution is to retranslate, not revise.

Whether we have been influenced by the source text or the original translator has consciously or unconsciously chosen to translate too literally, our job as self-revisers and editors is clear. We must identify the problematic translations and correct them. A task easier said than done. How can we be sure that our revised translations or edits are truly idiomatic, that they reflect the way the target language is actually written?

If you are a sensitive native speaker of the target language, this should not present too great a problem. Your ear should immediately pick up the errors. So if this is your background and you have never seen the source text, you may be safe from any undue influence. Even if you are a translator with a similar background, waiting a few days before self-revising may be enough to allow you to escape the influence of the source text. If this is your case, just be sure not to begin with a comparative rereading of the source text and the translation to check for accuracy and completeness, as this will refamiliarize you with the source text. Instead, begin with your unilingual rereading of the translation for correct English language and style.

However, for those of us who lack the time to wait and self-revise, who have been compromised by over-familiarity with the source language or who have resided for many years in countries where our mother tongue is not spoken (and the source language is), the situation is more complex. We may be genuinely uncertain whether or not an expression is idiomatic English or an unfortunate calque of the source language. Confusing the matter, some calques, which begin as overly literal translations, enter the target language. So maybe we should refer to ראש הממשלה as the “head of government”, and not as the “prime minister”. Maybe the

French “progressif” and English “progressive” or the Hebrew קניון and the English “canyon” are not “false friends”. Maybe the phrase cited in the title (and actually published in an English-language newspaper) “this is the situation and what are the chances it will ever change” is idiomatic English, but maybe it is an agonizingly literal translation of ומה הסיכויים שהיא ישתנה.

To maximize the number of errors we catch, knowledge is a prerequisite. We must be aware of the types of errors that typically creep into translations of the language pair we are dealing with. In terms of idiomaticity, we must be aware of typical errors of collocation, colligation, phrasal verbs, set phrases, clichés, expressions, and so forth that are “not the way we say things in the target language”. Having this foreknowledge, we are more likely to notice errors. However, to bolster our efforts, it is crucial to develop a procedure designed to catch these errors. Writing a checklist of these types of errors, and devoting one read through of the translation to finding them is integral to such a process.

Once we have identified possible errors, in order to deal with our suspicions, a good dictionary can often provide some support as can combinatory dictionaries, such as The BBI Dictionary of English Word Combinations, which provides the correct grammatical and lexical combinations for tens of thousands of words. Likewise, Wood’s English Prepositional Idioms (1967) or Cowie and Mackin’s Oxford Dictionary of Current Idiomatic English (1975) teach the correct collocation of prepositions. Collins Cobuild English Dictionary, which provides a full-sentence example with each word, demonstrates certain appropriate syntactic structures. Online you can try the Cobuild Concordance and Collocations Sampler at www.collins.co.uk/corpus/CorpusSearch.aspx. This will enable you to search an enormous corpus of texts for validation of the syntax used in the text you are revising or editing.

Sometimes, in the Internet age, Google may help as a quick search for an idiomatic word or word combination can immediately supply many examples; however, care must be taken to exclude those examples not written by native English speakers or found in other suspect texts, such as translations. Furthermore, certain idiomatic combinations may be appropriate in a particular genre, in a particular form of discourse – oral versus written, or for a particular English-speaking audience, say South Africans.

By refining your Google search you can eliminate some of these concerns. If you are searching for a particular genre or form of discourse, you can limit your search to a database dedicated to the field or written for a particular audience (theoretically, you can even create your own searchable corpus with the right computer skills). Alternatively, if, for instance, you require an academic level of writing in a particular field you can search for suitable books using Google Books. In the case of a particular English-speaking target audience, say South Africans, you can add “site:za” to the end of your search string to search only South African websites.

If you are translating in a field and language pair where many official translations and their source texts exist online you can search for these in, for instance, government or commercial websites that contain both versions. If your field is Canadian law, you may have a particularly easy time as you can go to www.tsrali.com to find an aligned French-and-English corpus of court

► (1) Many of the ideas and suggestions proffered in this article owe a debt to Brian Mossop’s masterful *Revising and Editing for Translators* (Manchester: St. Jerome Publishing, 2007). Particularly useful for further reading are chapter ten, devoted to revision parameters; chapter twelve, devoted to revision procedures; and the references cited at the end of the book.

TARGIMA

Special Issue

ITA Annual Conference

February 8-10 2010

The ITA wishes to thank the authors who contributed their materials for this special issue, Sarah Yarkoni for her marketing and editing efforts, Yafit Alsheich and Hagit Rozanes for their help, and Inga Michaeli and Ester Halac for editing the materials.

Contents

Corpora Translators Can Use in Their Daily Practice

Ana Julia Perrotti-Garcia · 3

Revising and Editing Translations: Must We Admit that “This Is the Situation and What Are the Chances It Will ever Change”?

Meshulam Gotlieb · 5

A New-Found Craze

Gila Blits· 6

Translators, Translation Companies and Professional Associations: Why Collaboration Benefits Everyone

Rina Ne’eman · 7

Seven Free Tools to a Make a Translator’s Day

Naomi Elbinger · 8

The 2010 Translator's Horoscope: Get Ready for More Change

Ilan Bloch· 9

ApSIC Xbench, the Small Search Macro that Became a Localization QA product

ApSIC· 11

Graphic Design: Graphically, Yona Kidron-Shalev

Corpora Translators Can Use in Their Daily Practice

by Ana Julia Perrotti-Garcia, LAEL PUCSP, Brazil
(drajulia@terra.com.br)

The Internet has undeniably major impact on translating practice. Libraries, bookshops, journals and news websites are just some of the sources translators can access, most of the time without having to pay anything or paying very little. On the one hand, we have plenty of material available online, but on the other hand, this material is heterogeneous, unclassified, and not always 100% reliable. Because of these shortcomings, a number of authors have started thinking about other ways to use the Internet.

A corpus is a collection of texts of written or spoken authentic language presented in electronic form, providing evidence of how language is used in real situations. They allow translators to glean accurate and natural terms and sentences in a given language. As we usually tell our students, grammar books can show you how a language **should be used**, but it is the corpus that will show you how speakers/writers **actually** express themselves in the real world.

By using the corpus, the translator can examine a word in detail by looking at all the different contexts in which it occurs. If you search for a word in a corpus, the results are generally displayed in lines of concordance, i.e., the search word, highlighted and (usually) in bold, surrounded by a fixed number of words/characters predefined by the program and/or the user. There are hundreds of language corpora available online. Bilingual and monolingual corpora are rich sources of authentic texts translators can use to search for vocabulary, phraseology and grammatical information. Some of these corpora can only be searched for online; others can even be downloaded to be used with specific computer programs (note that the latter usually require more knowledge, while the former are more user friendly, so that if you are novice in the field of corpus research, choose the searchable online corpora as you will not need to worry about downloading a corpus search engine, such as a concordancer).

There are corpora in different languages and variants (such as Portuguese, Russian, Ancient Greek, Esperanto, Ebonics), and there are also corpora whose texts were selected by theme or subject (for instance medical corpora, children’s language corpora, suicide notes corpora).

Some English corpora available:

The **British National Corpus** (BNC, available at <http://www.natcorp.ox.ac.uk/>) is a 100-million-word collection of samples of written and spoken language, designed to be a representative cross-section of current British English. The written part of the BNC (90%) includes extracts from British newspapers, specialist periodicals and journals on a wide range of subjects and areas, academic works and popular fiction, letters and memoranda, and academic essays. The spoken part (10%) consists of transcriptions of informal conversations and spoken language collected in different contexts, ranging from formal business or government meetings to radio shows.

TARGIMA
Special Issue
ITA Annual Conference
February 8-10 2010

**FEB
2010**

