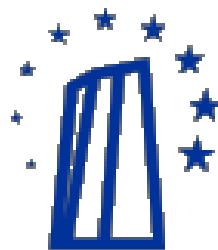




ПРОУЧВАНЕ НА ЕЦБ ЗА МАШИННИЯ ПРЕВОД



Жанета Шинкова
DGC/LGS/MTE3



Преводаческите услуги в ЕЦБ

- Отдел „Езикови услуги“ е част от ГД „Комуникации“.
- МТЕ/ЕТЕ/LTI
- Задачите за превод идват от различни структурни звена: комуникации, икономика, парична политика, банков надзор, банкноти, статистика, платежни системи, ЕССР и др.
- Правните актове се превеждат от юрист-лингвистите в ГД „Правни услуги“.



Какво е специфично за нашите текстове?

- От популярни до много технически; могат да бъдат политически или пазарно чувствителни.
- Някои са с висока степен на конфиденциалност.
- Дължината им варира в широки граници.
- Регистърът варира съществено според аудиторията, за която са предназначени.
- Не всички документи се превеждат на всички езици.
- ЕНМ промени значително обема и разпределението на работата между езиците.



Използва ли се в ЕЦБ машинен превод?

- В момента отдел „Езикови услуги“ използва само човешки превод, подпомаган от САТ и терминологични инструменти.
- Експертите от ЕНМ имат достъп до MT@ЕС като средство за бърз ориентировъчен преглед на документи, за да решат дали да ни ги изпратят за превод.
- Достъпът до невромашинния инструмент е блокиран, докато се проверяват аспекти на сигурността.



Какво искахме да установим с проучването?

Целта на проучването за машинния превод беше да проверим качеството на няколко инструмента за МП при различни текстове, от EN към 22 официални езика и от DE, ES, FR, IT, и NL към EN.

МП е деликатна тема, към която се опитахме да подходим без предразсъдъци, прагматично и отворено, отчитайки спецификата на превода в ЕЦБ.



Организация на проучването

- МТЕ: Набор от 5 текста, едни и същи за всички езици, представителни за различни трудности: техническа сложност, специализирана терминология, дълги изречения с тежък синтаксис, метафоричност, некачествен изходен текст
- LEG: Отделен набор от 3 правни текста
- ETE: От 5 основни езика към английски



Методология

- Инструменти: MT@EC, eTranslate, Google Translate (само за сравнение), DeepL (за езиците, за които е наличен, за BG не е)
- Сляпо проучване – преводачите не знаят кой инструмент оценяват
- Текстове с дължина до към 20 сегмента
- Категории грешки: терминология, грешен превод, граматика, пропуск, стил, [безсмислица]
- Цялостно впечатление: 4 категории: „за публикуване“, „за редакция“, „за ориентация“ и „неизползваемо“ ('fit for publication', 'fit for post-editing', 'fit for gisting', 'unusable')



Текст 1: обяснителна статия, **популярен език, широка аудитория**

"In a currency union, money travels. As euro banknotes travel freely across borders, at any one time the number of banknotes in circulation registered at each national central bank can vary enormously. For example, when German tourists seeking a Latin adventure take euro banknotes issued by the Deutsche Bundesbank with them to Spain, these notes are finally lodged at the Banco de España. As banknote issuance is an entirely demand-driven process, this means that the Banco de España would now need to put fewer notes into circulation because it is benefiting from the notes crossing the Pyrenees."



(Текст 1)

	Статистически	Невромашинен	Google Translate
Силни страни	Някои сполучливи термини: „сеньораж“, „банкноти в обращение“, „емитиране на банкноти“	Добра граматика, четете се гладко, някои от проблемите могат лесно да се поправят (напр. неправилни предлози)	Добра граматика, малко тромаво звучене
Слаби страни	Граматически грешки и пропуснати думи, водещи до голям брой неправилни преводи	Някои думи остават на английски; липсващи думи, водещи до неправилен превод; може да звучи подвеждащо гладко – и да е пълна безсмислица	Доста грешки в превода и терминологията
Безсмислици	6/12	5/12	0/12
As banknote issuance is an entirely demand-driven process, this means that the Banco de España would now need to put fewer notes into circulation because it is benefiting from the notes crossing the Pyrenees.	Емитирането на банкноти в изцяло е процес, това означава, че търсенето на Banco de España ще трябва да създадат по-малко отбелязва в обращение, тъй като то се ползва от отбелязва пресичане на Пиренеите.	Тъй като емитирането на банкноти е процес, основан изцяло на търсенето, това означава, че Banco de España сега ще трябва да приведе по-малко бележки в обращение, тъй като се възползва от обяснителните бележки, преминаващи през Пиренеите.	Тъй като емитирането на банкноти е процес изцяло насочен към търсенето, това означава, че Banco de España сега ще трябва да пусне по-малко банкноти в обращение, защото се възползва от нотите, пресичащи Пиренеите.



Текст 2: Лекция на Ив Мерш (метафоричен език)

"European folklore warns of the will-o'-the-wisp, a malignant creature that dwelt in marshes. It would appear as a light in the distance, which a traveler would mistake for houses. As they reached the place where they thought the light was, it would move further ahead, drawing them deeper into the marsh to their untimely death and a watery grave. The most recent beguiling wisps are named variously "cryptocurrencies" – to denote the use of cryptographic methods and technology – or "virtual currencies" (VCs) – to denote their lack of legal recognition."



(Текст 2)

	Статистически	Невромашинен	Google Translate
Силни страни	Прилична терминология: „тесен паричен агрегат“, „валутни зони“, „виртуални валути“. Правилно използване на конвенции като кавички, тирета, %	Добра граматика. Прости и по-технически изречения се получават като цяло добре. Правилно използване на конвенции като кавички, тирета, %	Отделни правилни термини: „тесен паричен агрегат“. Граматиката е на места добра, на други – не.
Слаби страни	Граматиката. Отделни думи остават на английски. Голям брой неправилни преводи.	Има изречения, които звучат гладко, но са пълна безсмислица. Голям брой неправилни преводи.	Много буквален превод, затова често няма общо с действителността. Нелогични и непоследователно използвани акроними.
Безсмислици	12/18	9/18	5/18
It would appear as a light in the distance, which a traveller would mistake for houses.	Изглежда в светлината на разстояние, която даден пътуващ би грешка за къщи.	Той би изглеждал на „светло“ разстояние, което пътуващото лице може да направи по погрешка в къщите си.	Това би изглеждало като светлина в далечината, която пътник щеше да греша за къщи.



Текст 3: Годишен доклад на ЕНМ 2017 – технически текст, специализирана терминология

"The ECB is constantly monitoring the sensitivity of banks' interest rate margins to interest rate changes. In the context of the low interest rate environment, which affects the profitability of the banking sector, the ECB decided to carry out a more in-depth assessment in 2017 of the strategies developed by banks to maintain the level of their interest margins under several scenarios. Consequently, in the first half of 2017, ECB Banking Supervision ran a "Sensitivity analysis of interest rate risk in the banking book (IRRBB) – Stress test 2017". A sample of 111 significant institutions (SIs) was assessed based on two complementary metrics: (i) changes in the banks' net interest income (NII) triggered by interest rate movements; and (ii) changes in the banks' economic value of equity (EVE)[3] (i.e. the present value of their banking book) triggered by interest rate movements."



(Текст 3)

	Статистически	Невромашинен	Google Translate
Силни страни	Добра терминология	Гладко, четете се добре, граматически правилно, добра терминология	Добра граматика
Слаби страни	Слаба граматика, проблемни акроними, голям брой неправилни преводи	Неправилни преводи, включително заради неправилни акроними	Много буквално и съответно тромаво; пълна безсмислица в някои сегменти
Безсмислици	13/20	4/20	3/20
Moreover, modelled durations of core deposits were found to be surprisingly long in some cases.	Освен това моделирани продължителност от основни депозити бяха сметени за изненадващо дълга в някои случаи.	Освен това е изненадващо, че по време на моделирането продължителността на основните депозити е изненадващо дълга в някои случаи.	Освен това в някои случаи е установено, че продължителността на моделиране на сърцевинните отлагания е изненадващо дълга.



Текст 4: Прогнози (технически, сложен синтаксис)

"Looking at the components, although HICP energy inflation is expected to strengthen in the coming months, reflecting the recent increases in oil prices and some upward base effects, from the end of 2018 it should decrease sharply, falling to rates around zero in mid-2019. The flat path of HICP energy prices from mid-2019 reflects the combination of a downward-sloping pattern of the oil price futures curve and upward effects from energy taxes. HICP food inflation is envisaged to pick up moderately over the projection horizon, supported by assumed increases in international food commodity prices and upward effects from increases in tobacco taxes."



(Текст 4)

	Статистически	Невромашинен	Google Translate
Силни страни	Простите изречения се получават правилно	Добра граматика, четете се гладко, като цяло добър превод	Като цяло приличен превод, отчасти добра терминология
Слаби страни	Голям брой неправилни преводи поради слаба граматика	Отделни граматически грешки; терминологията е частично сполучлива; на места правилен, но тромав превод	Грешни акроними; на места неправилен превод поради лоша граматика, тромаво
Безсмислици	4/11	0/11	0/11
The flat path of HICP energy prices from mid-2019 reflects the combination of a downward-sloping pattern of the oil price futures curve and upward effects from energy taxes.	Единната пътя на ХИПЦ цени на енергията от средата на 2019 г. отразява комбинацията от downward-sloping модел на цената на петрола фючърси крива и възходяща последици от данъците върху енергията.	Плоският път на цените на енергията по ХИПЦ от средата на 2019 г. отразява комбинацията от наклонена тенденция на спад на цените на петрола и възходящо въздействие от енергийните данъци.	Плоският пътеки на цените на енергията в ХИПЦ от средата на 2019 г. отразява съчетанието от низходящ наклон на кривата на фючърсите на цените на петрола и възходящите ефекти от енергийните данъци.



Текст 5: Слаб изходен текст

"As the model changes deal with both the concept for the definition of default and the validation concept, specific models (design and estimation), for which the respective permissions were already granted, are not affected by the model changes. [...] The ECB would like to stress the need to remove existing impediments, notably related to governance issues, as far as possible ahead of a stressed scenario to increase its resilience."



(Текст 5)

	Статистически	Невромашинен	Google Translate
Силни страни	Интересно е, че е хванал и оправил грешка в оригинала; добра терминология	Правилно тълкуване на акронимите; граматически правилен превод.	Прилична граматика
Слаби страни	Отделни думи остават на английски; слаба граматика	Смесени резултати – някои много добри, други много слаби сегменти. Изпуснати са важни думи, заради които смисълът се изкривява.	Малък брой сегменти са пълна безсмислица, но много голям брой неправилни преводи. Отделни думи остават на английски.
Безсмислици	7/12	5/12	3/12
Due to the economic trend also in ABC numbers and volumes of loans in restructuring and workout have declined for years.	Поради икономическата тенденция и в ABC и броя на кредитите, реструктуриране и реструктуриране намаляха в продължение на години.	Поради икономическата тенденция също и в броя на ABC и обемите на кредитите за реструктуриране и реструктуриране, реструктурирането и реструктурирането са намалели за годините.	Поради икономическата тенденция и в броя и обема на ABC кредити в реструктуриране и тренировка са намалели в продължение на години.



Заклучения (1)

- МП от английски към български все още е с много непоследователни резултати.
- Техническите текстове, включително с трудна терминология, всъщност се получават по-добре от „популярните“. Редовни публикации се получават отчетливо по-добре. Най-голямо затруднение е метафоричният език.
- Статистическият инструмент знае правилните думи, но често не успява да предаде смисъла, главно поради граматическо несъвършенство. Може би това е причината да работи по-добре в посока от BG към EN.



Заклучения (2)

- Google Translate се справя добре с популярен текст. Качеството му е напреднало (може би поради голямата натрупана маса текстове?). Много буквален превод. Затрудняват го технически сложните текстове и терминологията.
- При невромашинния инструмент граматиката е добра, текстът тече гладко (но това може да е подвеждащо – преводът да е пълна безсмислица). Сполучливите сегменти са наистина сполучливи. Грешните са зрелищно грешни.



Проучването продължава

- Нашето проучване още не е завършено. Все още се анализират резултатите – както в „Езикови услуги“, така и в правния отдел.
- Най-интересното предстои – сравнението между резултатите при различните езици.
- Предвижда се втори етап, в който да проверим приложимостта (включително в техническо отношение) на интегрирането на МП в инструментариума наред с традиционните памети.