Multilingual Environmental Terminologies

The GEneral Multilingual Environmental Thesaurus (GEMET)1 has been developed as an indexing, retrieval and control tool for the European Environment Agency (EEA), Copenhagen, its network of Member Countries (Eionet) and other environmental stakeholders since the beginning of 1996 (Budin, 2007). During the following five years, development work was rather intense and has been undertaken by partners of the EEA in Germany, Austria, Italy, Spain and Sweden as well as with the Environment Program of the United Nations (UNEP). The basic idea for the development of GEMET was to use the best of the available multilingual thesauri as a starting point, in order to save time, energy and funds. GEMET was conceived as a "general" thesaurus, aimed to define a common general language, a core of general terminology for the environment. Specific thesauri and descriptor systems (e.g. on Nature Conservation, on Wastes, on Energy, etc.) have been excluded from the first step of development of the thesaurus and have been taken into account only for their structure and upperlevel terminology. A set of existing national – partially multilingual – thesauri from Germany, Italy, The Netherlands, Spain and France as well as vocabulary from EEA assessments and from the EU Commissions Eurovoc thesaurus2 were merged (Steinberger, Pouliquen and Hagman, 2002).

The merging has been performed both on conceptual and formal basis. Coinciding concepts in the different thesauri have been identified and scored. Like in other multilingual thesauri, for example, UNEP's Infoterra EnVoc, a neutral alphanumerical notation allows the identification of a concept independently on the user's language. The resulting 6,562 terms have been arranged in a classification scheme made of 3 super-groups, 30 groups plus 5 accessory, instrumental groups. Each descriptor has been arranged in a hierarchical structure headed by a Top Term. The level of poly-hierarchy, that is, the allocation of a descriptor to more than one group, has been kept to a minimum. Further, to allow a thematic retrieval of terms thematically related but scattered in different groups, a set of 40 themes have been agreed upon with the EEA and each descriptor has been assigned to as many themes as necessary. Thus, the user can access the thesaurus through the group-hierarchical list, through the thematic list or through the alphabetical list. As a complement to the hierarchical "vertical" relations, an exhaustive series of strong "horizontal" relations between terms have been introduced.

The first published version of GEMET presented 5.298 descriptors, including 109 Top Terms, and 1.264 synonyms in English. This version contained translations funded through different sources into several other EU languages but also others like Russian or the Basque regional language. The GEMET project has been a driver for cooperation with the United States Environment Protection Agency (EPA) since 1999. The development of a "Common Global Environmental Vocabulary" – jointly between EEA, UNEP and EPA was officially announced in February 2000.3 A "Terminology Project Report"4 described the links made with the EPA's Terminology Reference System (TRS). Spanish terms in GEMET were of special interest for the American community and thus included into the TRS.

GEMET was shared in different ways over the years. A limited amount of volumes was printed, a PC (Windows) version was developed as well as a web version existed by 2001. As one of the first vocabularies, GEMET content was made available in the format of the Simple Knowledge Organisation System (SKOS)5 (Miles and Pérez-Agüera, 2007) and since the web version had an API it could be widely accessed and connected to. Later on, queries based on a semantic query language for databases (SPARQL)6 were added. It was also one of the first sources included into the Linked Open Data (LOD) cloud. After 2001, GEMET was hosted directly by EEA and a dedicated website was developed. In the years following, maintenance focussed on translations. Bulgarian, Czech, Estonian, Russian and Slovenian have been added in 2001.

The addition of Arabic (2009) and Chinese (2010) marked another significant step in internationalisation. Further (regional) languages added in 2012 were Catalan, Slovak, Croatian and Ukrainian. In the same year and for the first time, a smaller set of 15 additional terms was inserted. Thanks to a project with European Neighbourhood countries, Armenian, Georgian and Azerbaijani were included in 2015 and it succeeded in 2017 to include with Icelandic the last European language not yet covered. For selected languages and based on stakeholders' interest, the English definitions were translated over the years. Overall, GEMET terms are today available in 37 languages and after a first substantial term update in 2018 with 100 newly added and translated terms (plus further 200 on a candidate list) it contains close to 5500 terms.

In an Internet survey done in 2000, a strong overall usage of GEMET was identified. Most of the initial thesauri but also other largely thematic thesauri and vocabularies were linked to GEMET through the API. This became more and more popular as the usage of the Internet was growing. It became also clear that usage went beyond data indexing – translation support was of a growing application area. No second study on how this developed over time has been done. The usage of controlled vocabularies to support website translations is an application area, EEA is currently considering.

On the occasion of this article, EEA ran an evaluation of the GoogleAnalytics statistic regarding GEMET usage. Overall, in the year until mid-June 2018, approximately 300,000 user visit were counted. Out of these visits, half are attributed to "organic" search, which includes machine access either through search engines or application programming interfaces (APIs). One quarter each originates to "direct" (a user enters the GEMET URL) and "referral" (another website links to GEMET) usages. Half of the users utilise the English language, which mirrors roughly the geographic distribution of the origination of the users (evaluated through internet protocol (IP) addresses). EEA considers this an excellent uptake of the GEMET service.

Looking at the development patterns of GEMET in the past 22 years, there have been technical and political reasons which set priorities and in particular triggered the inclusions of languages. The technical demand was steered by the need of EEA to get an overview of the available environmental data in EU Member States. The metadata related to environmental datasets with European relevance was collected and registered in a metadata catalogue. Controlled vocabularies were used to index this metadata. In the countries, some of the thesauri supported registration functions in their libraries of environmental literature.

Political reasons ranged from so different motivation as the wish to belong to the "European environmental family" and the interest to document national or regional identity in a world with a level of English language dominance to the interest to cooperate stepwise on technical environmental matters before providing eventually later access to environmental data. A main challenge in GEMET development has been governance and content update cost. The governance aspect is related to identifying and running the right processes to identify update needs. Usage by EEA Member Countries has always been diverse – some were rather interested – others did not see a need for a common vocabulary. A good buy-in from the countries is needed to initially agree on an update and later on motivate quality assurance. The latter is one of the reasons why a first content update was only done in 2012, including few key terms, which came into the environmental debate and were missing in GEMET. The strongest example here is "climate change" and its narrow terms.

In 2016, EEA identified the need to start a more substantial content update. Since there is no obvious methodology on how to do that, after some discussion, the following approach was agreed: Circa ten key EEA publications including the 2015 report on the state of environment in Europe as well as the vocabulary used on the EEA website was submitted to the Translation Centre for the European Union (CdT). They did a text analysis resulting into 300 most popular terms and included definitions for them. Out of those, EEA identified 100 terms in 23 languages and submitted them to the contact points in all 28 member states for quality control. After a second round of internal quality control, results were included into a GEMET version 4.1.0. A shortcoming is that the new additions are not available in the remaining 14 non-EU languages.

Considering the effort to systematically involve stakeholder as described above, alternative approaches to collect terms for updating an environmental core vocabulary have been explored shortly after 2000 under the Wikimedia Commons7 related Wiktionary project. The idea was that users would suggest new terms in a process, which would have needed to be moderated but this did not happen to any useful numbers. Maybe a social media-based approach would be more promising these days on the other hand that could conflict with EEA or member states priorities.

A recurring task has been to define and apply both communalities and the distinction between an environmental core vocabulary like GEMET and domain vocabularies. In the EEA context, these are currently controlled vocabularies for biodiversity, water and climate change adaptation. A relationship can be organised through building thesauri federations. The way to do that technically and methodologically still needs to be identified. One example for orientation to lead out such work is the sematic alignment between EuroVoc – the earlier-mentioned authoritative vocabulary of the European Union – and GEMET published in the EU Open Data Portal8 through a linked open data (LOD) approach.

A federation solution may be also built using the current implementation of relations in GEMET based on close or full (SKOS) matches to selected national thesauri and to Agrovoc9 – the well-elaborated thesaurus of the UN Food and Agriculture Organisation (FAO) or to EuroVoc. International thesauri initiatives are welcome to keep using GEMET content through the API or through the vocabulary published in linked open data. Given EEA's resource constraints, it will not be possible to systematically and directly liaise with other communities

outside its own Eionet network for content updates or specific technical solutions. Since EEA at the same time is committed to provide reliable web services and has an open data policy, access to GEMET is possible in several user-friendly ways. These measures assure global usage by interested institutions into the future.

The UN Sustainable Development Goals provide abstract, overarching governance frameworks (Kanie and Biermann, 2017) which requires important local adaptation and implementation through cross-sectoral collaboration. In the study of the GEMET database, we noticed that the Chinese translations collected in the database developed two decades ago, now only covers a fraction of the large, increasing number of sustainability-related expressions appeared in Chinese official publications and mainstream. With the introduction of translated environmental resources, new locally and culturally adapted sustainability expressions and terms have been created and used extensively in official and everyday Chinese resources. To offer an overall analysis of the growing environmental discourse in China, we developed the Chinese-English Environmental Translation Terminology (CEETT) which aims at supplementing existing multilingual environmental resources such as GEMET.

The CEETT homepage presents a complete list of all the bilingual sustainability terminologies. Term entries are arranged in an alphabetic order to facilitate browsing and look-up. The system is to a large extent a search engine working on restricted data sources. Users can enter any intended search terms, be it English or Chinese. The system will return a list of bilingual terms containing the usersupplied search string. Fuzzy query is set by default in our system. For instance, when 生产 is used as a search word, such bilingual terms as 本地生产 (Local production), 错峰生产 (Staggered peak production), 能源生产革命 (Energy production revolution), 清洁生产 (Clean production), 全要素生产率 (Total factor productivity), and 生产活动 (Production activities) are returned. The fuzzy query mechanism locates any entries in the database with initial, middle, final part or exact match of the text string 生产. It is the same case with English searches (e.g. “production” as the query word).

Fuzzy search renders CEETT users more relevant terms, which has important applications in environmental translation studies. Terminologies are one of the major hurdles in professional translation and usually not semantically transparent to novice translators. Therefore, explanatory notes (e.g. Domains of Knowledge, Sustainability Development Goals, SDG Implementation Analysis Framework, Sources of Notes, Excerpts of Chinese Texts, and Notes) are available to each entry following the hyperlinks of individual hits. For instance, the notes for 全要素生产率 (total factor productivity) go like 指一个系统的总产出 量与全部生产要素真实投入量之比，测算公式为：全要素生产率=产出总量/ 全部资源投入量 (total factor productivity refers to the ratio of the total output of a system to the actual input of all production factors. The calculation formula is: total factor productivity = total output / total resource input load).

Authorised users can sign in the Admin Centre with credentials to edit the existing terminological expressions. Both the entry proper as well as its explanatory metadata can be modified whenever necessary. New entries can always be added at the Entry Maintenance page of the Admin Centre. The project investigators can add new users and assign relevant roles for them to either contribute new terminological entries or maintain the entire database. Our next

planned expansion of the system is to link all the terminologies to parallel concordance lines of environmental corpora or other general-purpose parallel corpora. More real-life usages will inevitably facilitate the mastery the technical terms as well as the idiomatic use in proper translational context. The basic structure of CEET resembles that of GEMET and the main difference between the two is the direction of translation. While GEMET contains multilingual terminologies translated from English as lingual franca to other languages including Chinese, CEETT encompasses a large and growing number of English translations of environmental expressions that have featured in Chinese official publications and mainstream media, which provide first-hand materials of the current discourse of environmental protection and sustainable development in China. The CEETT was compiled in JavaWeb. The bilingual terminologies were saved as an Excel spreadsheet before they were fed into the web application. The data sheet was then converted to database format to be compatible with the JavaWeb programme.

This list contains some basic features of the Chinese-English environmental translation database: 1 Translation Mode (how the Chinese sustainability expression was created): literal translation of English terminology (ref. GEMET term base); or locallycreated expression; or a mix of both 2 Word Length in Character: 3, 4, 5, or 6, which is an indication of conceptual complexity or cognitive load – long expressions would need to be adapted for public environmental education 3 Domains of Knowledge (comparable to the thematic classification in GEMET) 4 Sustainability Development Goals (SDGs) 5 SDG Implementation Analysis Framework (we created this open-ending list to facilitate policy analysis): Problems, Targets, Standards, Principles, Approaches, Methods, Actions, Actors, Resources 6 Sources of Notes: explanations of the semantic meaning of Chinese sustainability-related expressions 7 Examples: examples taken directly from Chinese official publications 8 Notes: any other notes. The ultimate goal of the database is to assess how local/national sustainability language/culture has been developing after 20 years of the translation and social dissemination of environmental policies (such as the creation of GEMET by the European Environmental Agency). They could be used as matching databases for multilingual databases like GEMET which contain translations from English into other languages.

In our study, we used both the GEMET and CEETT databases to study the translation of sustainable terminologies in Chinese official publications and the media, which help us understand the patterns of the growth of sustainable discourse in the country. In our study of original Chinese materials, all data were collated from the Factiva database (Dow Jones 2017), a global news database consisting of a wide range of licensed and free publications in numerous languages including Chinese. The database includes a large range of licensed digital materials published by governmental, industrial and business sources in different countries in their original languages since the mid-1990s. Based on the exploration of the digital Chinese publications, four large groups of terms were extracted which are closely related to sustainable living and citizens' social responsibilities to protect their living and working environments. The four large word categories encompass translations from original English terms and a number of locally-designed expressions representing useful efforts to adapt abstract sustainability principles and goals into concrete and specific actions and behaviours. The four word categories highlighted are responsible behaviours; green ethics and social responsibilities for sustainability; green and sustainable living environments and lastly, green or sustainable transport and travel options. These four sets of sustainability terminologies were used to query

the database and extract frequency data of publications on particular topics and dimensions of promoting green living style and the public awareness of sustainability. It is hypothesised that the acceptance and circulation of these translated terminologies by different social agencies may have an impact of the sustainability discourse in China.

The first word category is defined sustainable consumption behaviour: 光盘行动 (guāngpán xíngdòng) (eat-it-up campaign,to avoid food waste); 理性消费 (lǐxìng xiāofèi) (rational consumption); 新文化消费 (xīn wénhuà xiāofèi) (consumption based on new cultures); 低碳消费 (dī tàn xiāofèi) (lowcarbon consumption) 低碳消费行为 (dī tàn xiāofèi xíngwéi) (low carbon consumption behavior); 理性购买行为(lǐxìng gòumǎi xíngwéi) (rational purchase behavior); 过度消费 (guòdù xiāofèi) (over consumption); 环境行为 (huánjìng xíngwéi) (environmental behavior); 环境教育 (huánjìng jiàoyù) (environmental education); 节水 (jié shuǐ) (water saving); 节能 (jiénéng) (energy saving); 节能量 (jié néngliàng) (energy saving); 节能降耗 (jiénéng jiànghào) (energy saving and consumption reduction); 家庭使用 (jiātíng shǐyòng) (family use); 节俭用餐 (jiéjiǎn yòngcān) (thrifty dining); 可持续消费 (kěchíxù xiāofèi) (sustainable consumption); 科学消费 (kēxué xiāo fèi) (scientific consumption); 垃圾分类 (lājī f ēnlèi) (waste classification); 铅回收 (qiān huíshōu) (lead recycling); 清洁利用 (qīngjié lìyòng) (clean utilisation); 省电 (shěng diàn) (electricity saving); 消费行为 (xiāofèi xíngwéi) (consumption behaviour); 自家消费 (zìjiā xiāofèi) (home consumption); 自给自足 (zìjǐ zìzú) (self-sufficient); 自我负担 (zìwǒ fùdān) (self-pay); 责任消费(zérèn xiāofèi) (responsible consumption); 汽车共享 (qìchē gòngxiǎng) (car sharing); and 智慧节能 (zhìhuì jiénéng) (smart energy saving).

The second word category refers to green ethics and social responsibilities for sustainability in Chinese: 碳贞操 (tàn zhēncāo) (carbon chastity); 低碳使命 (dī tàn shǐmìng) (low-carbon mission); 生态文明 (shēng tài wén míng) (ecological civilisation); 生态力量 (shēngtài lìliàng) (ecological power); 社会责任 (shèhuì zérèn) (social responsibility); 社会责任感 (shèhuì zérèngǎn) (social responsibility); 道德选择 (dàodé xuǎnzé) (moral choice); 环境责任 (huánjìng zérèn) (environmental responsibilities); 环保意识 (huánbǎo yìshí) (awareness of environmental protection); 环境伦理 (huánjìng lúnlǐ) (environmental ethics) 环境贡献 (huánjìng gòngxiàn) (environmental contribution); 价值观 (jiàzhíguān) (values); 价值取向 (jiàzhí qǔxiàng) (value orientation); 生产者延伸责任 (shēngchǎn zhě yánshēn zérèn) (extended producer responsibility); 企业社会责任 (qǐyè shèhuì zérèn) (corporate social responsibility); 社会公共利益 (shèhuì gōnggòng lìyì) (social public interests); 诚信经营 (chéngxìn jıngyíng) (business integrity); 消费者态度 (xiāofèi zhě tàidù)(consumer attitude) and 意识改革 (yìshí gǎigé) (awareness reform).

The third word category is green living environment which broadly includes built environments, communities, schools, future city design and sustainable lifestyle: 百年住宅 (bǎinián zhùzhái) (centennial residence, with energy-efficient and sustainable design); 低碳校园 (dī tàn xiàoyuán) (low carbon campus); 低碳乡村 (dī tàn xiāngcūn) (low-carbon village) 低碳社区

(dī tàn shèqū) (lowcarbon community) 低碳家庭 (dī tàn jiātíng) (low-carbon family) 绿色生活 (lüsè shēnghuó) (green life) 零碳生活 (líng tàn shēnghuó) (zero-carbon life) 绿色校园 (lüsè xiàoyuán) (green campus) 绿色设施 (lüsè shèshī) (green facilities); 生态村 (shēngtài cūn)(eco-village); 生态城区 (shēngtài chéngqū)(eco-city); 生态生活 (shēngtài shēnghuó)(eco-life); 生态住宅 (shēngtài zhùzhái) (ecological residence); 生态城市 (shēngtài chéngshì)(eco-cities); 低碳城市 (dī tàn chéngshì) (low-carbon cities) 低碳建筑 (dī tàn jiànzhú) (low-carbon building); 低碳生活 (dī tàn shēnghuó)(low-carbon life); 低碳环境 (dī tàn huánjìng) (low-carbon environment); 海绵城市 (hǎimián chéngshì) (sponge city) 可再生能源建筑 (kě zàishēng néngyuán jiànzhù)(renewable energy building); 绿色建筑(lüsè jiànzhù) (green building); 湿地公园 (shīdì gōngyuán) (wetland park); 消费生活 (xiāofèi shēnghuó) (consumption life); 永续家园 (yǒng xù jiāyuán) (sustainable home); 智慧家庭 (zhìhuì jiātíng) (smart families); 智慧城市(zhìhuì chéngshì) (smart cities); 智能家居 (zhìnéng jiājū) (smart homes); 城市环境 (chéngshì huánjìng) (urban environment); 垂直绿化 (chuízhí lühuà) (vertical greening); 持续环境 (chíxù huánjìng)(sustainable environment); 无废城市 (wú fèi chéngshì) (wasteless city); 智慧社区 (zhìhuì shèqū) (smart communities); 智能城市 (zhìnéng chéngshì)(intelligent cities); and 智能生活 (zhìnéng shēnghuó) (intelligent life).

The last word category studied is green transport and travel options which include high-frequency words such as 共享单车 (gòngxiǎng dānchē) (shared bike); 绿色交通 (lüsè jiāotōng) (green traffic); 零碳交通 (líng tàn jiāotōng) (zero carbon transportation); 绿色快递 (lüsè kuàidì) (green express); 智能物流 (zhìnéng wùliú) (smart logistics); 多式联运 (duō shì liányùn) (multimodal transport); 低碳交通 (dī tàn jiāotōng) (low-carbon transportation); 慢行系统 (màn xíng xìtǒng) (slow transportation system); 智能快递 (zhìnéng kuàidì)(smart express); 磁悬浮 (cíxuánfú) (Maglev); 磁悬浮列车 (cíxuánfú lièchē) (Maglev trains); 无碳交通 (wú tàn jiāotōng) (carbon-free traffic); 运输路径 (yùnshū lùjìng) (transport paths); 自动驾驶 (zìdòng jiàshǐ) (autopilot); 智慧交通 (zhìhuì jiāotōng)(smart transport); 低碳出行 (dī tàn chūxíng) (low-carbon travel); 零碳出行 (líng tàn chūxíng) (zero-carbon travel) and 微旅行 (wēi lü0xíng) (mini trips or short, low-cost trips).