On the Rationale of Trigger Word and Event Classification: A Case Study

Song Miaomiao

Peking University No.5, Yiheyuan Road, Haidian District, Beijing, China scarlettsun6@stu.pku.edu.cn

ABSTRACT. The lack of a consistent standard in the choice of trigger word and event types in event extraction has reduced the performance of language models and made it hard to evaluate the final results, thus hindering the further development of event extraction and classification research. To solve this problem, this preliminary study proposed a new set of event types for science and technology policy texts based on the ACE 2005 taxonomy. By comparing and analyzing the English and Chinese texts in this field, it is indicated that although ACE 2005 taxonomy provides a basic schema for researchers to consult, it needs a lot of adaptation when dealing with the data under specific topics. Moreover, the linguistic differences between Chinese and English leads to different methods in trigger identification, which needs more consideration in future research.

Keywords: trigger words, event extraction, event classification

1. **Introduction**. Digital technologies have advanced more rapidly than any innovation in our history. Today, with more and more text resources on the Internet, information extraction (IE) has been highlighted in natural language processing. It addresses the automated retrieval of structured information related to a selected topic from a body or bodies of text to assist a computer in understanding natural language [1]. Tasks in IE usually includes named entity recognition (NER), relation extraction (RE), co-reference resolution (CO), event extraction (EE) and other related tasks like delineating descriptive information from detected text [2].

As a key form of information representation, event plays a major role in bridging original texts and the processed structured information. According to ACE event description, an event "is a specific occurrence involving participants", "is something that happens" and can "frequently be described as a change of state" [3].

As the basic unit in text to describe an event, a sentence usually contains event arguments and event triggers. The former links with named entity and domain knowledge, and the latter – the word that "most clearly expresses the occurrence of an event", has received a lot attention in IE research. Due to the syntactic differences between language, the way in identifying trigger words has not reached an agreement, especially in Chinese event extraction. Meanwhile, with more and more approaches for EE that have been proposed, the benchmark to assess the performance of each model is

still debatable due to the lack of rationale. Researchers use different methods to label trigger words and use the same event type to classify events in different domains, making the results unreliable. Therefore, this study mainly addresses these issues. Through a case study, we first analyze the difficulties in trigger labeling caused by syntactic differences between Chinese and English. Then we propose a new set of event types that are specific to science and technology policy texts (or sci-tech policy texts, which will be used in the rest of this paper). This preliminary work can be a reference for future studies in event extraction and event classification.

2. Literature Review.

2.1 Overview of Event Extraction Research Categories. As mentioned previously, both language type and research domain can affect trigger labeling and event extraction. Generally, EE can be studied on sentence level or on paragraph level (based on text analysis granularity). It can also be categorized according to different language type, such as English event extraction, Chinese event extraction, Japanese event extraction and so forth. Judging by the content of data, event extraction contains open domain event extraction and closed domain event extraction. The former aims at detecting events from common text and usually clustering similar events via extracted event keywords. The latter, closed domain event extraction, often requires "predefined event schema to detect and extract desired event types from text" [4].

Today, researchers focus more on paragraph-level event extraction and open domain event extraction, which usually take machine learning or deep learning as research methods. At the same time, to avoid the difficulties in trigger identification caused by lack of standards, trigger free event detection has also become a tendency [5-6]. However, current open domain research lacks the interpretability of models. Trigger free approaches and deep learning approaches seem to avoid those difficulties. yet the semantic information of trigger words - the key point to address events, have also been lost. On open domain research, the best performance on English event classification/trigger labeling on the ACE 2005 dataset is the Recurrent Neural Network model of Li et al. [7], whose F1 value is 75.6%. On Chinese event classification/trigger labeling the number is 70.2%, done by Li et al. [8]. The results of closed domain event extraction using machine learning or deep learning methods are even more discouraging. Liu et al. [9] proposed a new learning paradigm of EE, casting it as a machine learning comprehension problem rather than a classification task. They used an unsupervised question generation process as well as a BERT- based question-answering process to retrieve answers as EE results, which give F1 score as 63.6% on argument extraction. Wadden et al. [10] examined the capabilities of DyGIE++, a unified, multitask framework for different IE tasks. Under limited resources, the F1 score on trigger classification task was only 69.4%. The authors themselves also pointed out that "due to the high sensitivity of both BERT finetuning and event detection to the choice of optimization hyperparameters - in particular, the trigger detector begins overfitting before the argument detector is finished training".

2.2. **Trigger Choice in Closed Domain Research.** Although ACE 2005 has already provided a model for the choice of trigger, which includes annotating complex examples, yet to EE in closed domain, there is no standard or reference to turn to when dealing with trigger words. This also makes the performance of different models less evaluable and reliable. Researchers usually takes the following two approaches to choose triggers in closed domain:

- (1) By manually built trigger lists. This is the most precise method and can give the best result. However, as Kilicoglu and Bergler [11] said, this dictionary-based approach is "naturally limited to the trigger expressions encountered in the training corpus and a small set of words obtained by expansion". Furthermore, due to the "one event type per word" constraint, contextual disambiguation of trigger word is not allowed.
- (2) By machine learning methods or machine learning combined with trigger list. Tian et al. took this combination approach and get F1 score 0.71 [12]. However, one problem of their study is that the data they use is Chinese Emergency Corpus (CEC), which only has 332 texts. Its event tagging system includes movement, emergency, statement, action, and so on. The relation between tagging and this closed domain is not strong enough. The reason behind this tagging (or classification) has not provided either.
- (3) By other statistical methods. Nowadays trigger-free event detection is an approach that directly overcomes the difficulties of trigger choice and reaches the final goal of event extraction and classification. Yet to avoid the problem at the beginning does not mean the problem can disappear. For example, the baseline of Deng et al.'s research uses TF-IDF method, taking the top 20 verbs and nouns as triggers [13]. Choosing triggers in this way cannot guarantee the reliability of new EE methods that this study proposed.

2.3 Syntactic Differences and Trigger Labeling. Syntactic differences between English and Chinese are quite prominent, which causes more difficulties in Chinese event extraction. ACE 2005 taxonomy has explained how to annotate event triggers. In many cases, English triggers are verbs in the part of the sentence (extent) that most directly describes the event. Sometimes, they are in the form of adjective or a past-participle. Examples in which the event is triggered by a noun or pronoun do exist, but they are quite rare in daily texts. Compared to the verb-in-nature characteristic of English language, Chinese is more like noun in nature. 77% of common English words have verb root, but Chinese characters are mostly composed of nouns, not to mention many sentences that have no verb at all [14]. Therefore, in preparing text data for EE, lemmatization is a major step used in English. But sentence segmentation and judging whether a word can be viewed as a trigger are all problems in Chinese that need to be solved. Chen and Li summarized four types of complex syntactic phenomena in Chinese texts [15]:

- (1) One sentence, multiple events.
 - (2) One word works as triggers for multiple events.
 - (3) One word can be viewed as multiple arguments in the same event.

(4) One word works as different argument in different events.

Deng et al. analyzed public security texts and gave such an example:

"我就问他一只'货'(冰毒)多少钱,他说 600元,我说拿一只给我,他就放了两小包 在老地方,我就从老地方将毒品拿回来了。"

Translation:

"I asked him how much one 'product' (drug) cost, he said 600 yuan. I said give me one. So, he put two small bags to the rendezvous, and I took them back."

In this example, only the two words "放" and "拿" coexist can one assure this sentence contains the event "drug trafficking". But the two words are quite common in daily use, which makes it harder for researcher to distinguish and extract the event [13].

To sum up, in closed domain EE research, trigger lists are still the most precise way to classify events. But most research today does not have a standard or agreement in trigger choice. Thus, the reliability of EE models' performance cannot be guaranteed, especially in specific closed domain where there is usually no finely tagged data. Moreover, the syntactic differences make trigger choice in other languages such as Chinese more difficult. Therefore, in the next section, we focus on these differences and provide solutions that may be helpful in processing Chinese text data. Later we use a case analysis to see which word can work as an event trigger in the field of sci-tech policy. This can be a reference for future closed domain EE research.

3. Chinese Triggers and English Triggers Comparison. Before making the comparison, we give a typical example of an event extracted from a sentence:

E.g., Top Gun 2 grossed nearly \$770 million overseas last year.							
Event Type/tag Trigger entity/argument entity/argum							
		Top Gun 2	Entity name				
Data and index	gross	last year	time				
		\$770	value				

 TABLE 1. Example of an event extracted from a sentence

3.1 Chinese Triggers. Below is an excerpt from Globaltechmap website:

据美国生物世界网6月5日报道 (report),美国3D植入物制造公司3DBio Therapeutics使用 (use) 3D细胞打印的活体组织植入物为一名右耳小耳畸形的患者重建 (rebuild) 了外耳。研究团 队先对小耳症患者的耳朵进行 (proceed) 活检,从中分离 (separate) 出软骨细胞,进行 (proceed) 体外细胞培养扩增,再使用 (use) 3DBio公司专有的创新3D打印技术制作 (make) 出 正常形状的耳朵,并交由 (hand in) 医生为患者进行 (proceed) 移植。该植入物由3D打印的胶 原蛋白水凝胶支架和患者的软骨细胞构成 (consist of),其结构与患者左耳的尺寸和形状相匹配 (match)。这种组织工程技术为治疗鼻子和脊柱缺损、乳腺癌术后的乳房重建以及器官移植奠 定了基础 (lay the foundation),在更广泛的再生医学领域中具有 (have) 巨大的应用前景。

Translation:

According to a report on June 5th by the bioworld website, 3DBio Therapeutics, an American 3D implant manufacturing company used 3D cell-printed living tissue implants to reconstruct the outer ear for a patient with microtia in the right ear. The research team first performed a biopsy on the ear of a patient with microtia, isolated chondrocytes from it, carried out in vitro cell culture and expansion, and then used 3D Bio's proprietary innovative 3D printing technology to produce a normal-shaped ear, and handed it over to the doctor to transplant to the patient. The implant is composed of a 3D printed collagen hydrogel scaffold and the patient's chondrocytes, and its structure matches the size and shape of the patient's left ear. This tissue engineering technology has laid the foundation for the treatment of nose and spine defects, breast reconstruction after breast cancer surgery, and organ transplantation, and has great application prospects in the broader field of regenerative medicine.

In the Chinese paragraph above, verbs and phrasal verbs are all marked as red. It is

obvious that apart from the first phrasal verb "据……报道" can be viewed as a trigger and tagged as "news reporting", none of the rest verbs can trigger an event, for they do not have any detailed semantic information that relates to certain fields. Thus, when handling Chinese texts, using merely verbs may cause great damage to the result. Here we suggest three ways to solve this problem:

First, do a part-of-speech tagging before trigger extraction. This belongs to word disambiguation process. Most Chinese words can work as verbs or as nouns without any morphological change, such as "重建" (rebuild or rebuilding), "分离" (separate or separation) and "构成" (compose or composition). It would be helpful if one uses trigger list to extract triggers and events, but how to distinguish words with specific domain information remains.

Another method is to elaborate the language model by including more items other than words. For example, the first upgraded model we can think of is that the coexistence of the verb with another word can determine event occurrence. In this way, trigger labeling and argument extraction can be processed at the same time. To some extent propagation of uncertainty can be avoided as well. The last solution works without trigger words. Different from previously introduced trigger-free analysis using machine learning, here it means abandoning detecting triggers in Chinese text and focusing on those entities and arguments in sentences. It may only work with Chinese texts, because it is common in Chinese language to use multiple nouns together to express meaning. This method of piling up *Yixiang* (imagery) needs no word. The semantic information of each item concatenated in sequence and their orders in a sentence can clearly represent an event. In event extraction, using triggers to define an event does not work well, but two or more entities appeared together is enough to justify that there must be something happen in this sentence, except those parallelism structure. Take the first sentence of the above paragraph as an instance:

TABLE 2. Using entities for	Chinese event detection
-----------------------------	-------------------------

E.g.美国3D植入物制造公司3DBio Therapeutics使用3D细胞打印的活体组织植入物为一名右耳小耳畸形的患者重建了外耳。				
label	content			
doctor/hospital	(entity) 美国3D植入物制造公司 3DBio Therapeutics			
patient	(entity) 一名右耳小耳畸形的患者			
method	(entity) 3D细胞打印的活体组织植入物			
diagnosis	(behavior) 重建			
body part	(entity) 外耳			

The whole sentence contains four entities and one behavior. If we exclude the verb/the behavior, we can also guess that there must be something occurred that related to health and medicine. Usually, two entities that belong to different categories are able to determine the occurrence of an event. This method might be helpful in future EE research. But it also needs more studies to justify its validity.

3.2 English Triggers. Compared with Chinese triggers, English triggers are much easier to judge. The excerpt below is from ABC News, titled "Some school districts bring back mask mandates to stem spread of COVID, flu and RSV":

Some school districts bring back mask mandates to stem spread of COVID, flu and RSV Two New Jersey districts announced the return of indoor masking. By Mary Kekatos December 31, 2022, 12:55 AM

Several schools are bringing back mask mandates to reduce the spread of COVID-19 and other respiratory illnesses.

In New Jersey, two school districts said students will be required to wear face coverings indoors upon returning from winter break.

Paterson Public Schools, which serves about 25,000 students from pre-kindergarten through 12th grade, said the new rule will go into effect starting Jan. 3.

"I know this is a relief to some, and a frustration to others. No matter what your position may be, I ask for your cooperation," Eileen Shafer, Paterson school district superintendent, wrote in a letter to parents and guardians. "Please continue to maintain universal masking throughout our buildings and we encourage you to take all other precautions against the spread of the COVID-19, RSV, flu virus including frequent hand washing, avoiding large gatherings, and staying home when sick."

All verbs are also marked red. In this example, "announced", "said", "required", "ask for", "wrote", "encourage" can all be labeled as "opinion and talks", and "reduce" can be labeled as "change of data". In English texts, not all verbs can be treated as triggers either, so the first two methods proposed for dealing with Chinese text can also be used here.

4. Trigger Labeling/Event Classification in English Texts. Extracting triggers is not the goal of research. The next step after trigger detection is often trigger labeling or event classification, since event type is usually determined by which category the trigger belongs to. In EE studies, trigger identification and labeling are usually done at the same time. In traditional trigger list methods, labels are predefined, such as the "opinion and talks" label mentioned above. This way of EE is not automatic and may be quite laborious when the research domain expands. Yet currently it is still the most precise and interpretable one in closed domain EE research. To figure out how to categorize triggers in a specific field will be useful for researchers when they choose baseline models in experiment. It will also benefit those automatic EE models by presenting insights of how we humans classify huge number of events. In this regard, we will first briefly introduce the set of event types in ACE 2005, then do a preliminary study, analyzing science policy texts and proposing a new set of event types to this domain.

4.1 Event types in ACE 2005. Most IE studies use ACE 2005 dataset to train and test their models. It is necessary to briefly introduce it here. ACE (Automatic Content Extraction) conference in 2005 provide examples of a particular set of event types and subtypes, which have been thoroughly discussed in the guideline [3]. The set of data types are shown below.

It is worthy to point out that the eight event types in ACE 2005 are not targeted at open domain research, although today it has been used in many open domain EE studies. The Linguistic Data Consortium [3]. mentioned that they will not tag all events, but

they are just "interested in annotating LIFE, MOVEMENT, TRANSACTION, BUSINESS, CONFLICT, CONTACT, PERSONNEL and JUSTICE Events and among these a particular set of subtypes". Therefore, this taxonomy is not the perfect one and can be further improved, especially for closed domain research.



FIGURE 1. ACE 2005 taxonomy.

4.2 A case study: event types for sci-tech policy texts. We choose science and technology policy as the closed domain for our research. Data in our corpus chiefly come from scientific journals, governmental official websites and think tank, such as Nature, U.S. Office of Science and Technology Policy [16] and Center for Strategy and International Studies [17]. More than 800,000 pieces of news have been included in the corpus. To analyze event types in detail, we choose 10 pieces of news each day from December 2022 to March, 2023.

Based on ACE 2005 taxonomy, we proposed a new set of event types targeting on our sci-tech policy corpus. During policy formation stage, more attention will be paid to the publication of policies, talks and comments of certain key figures as well as important events happened at that time. These have all been considered in classifying sci-tech policy events. Below is the taxonomy and trigger example for science policy texts. The 12 types have no subtype, since science policy is already a small research



field and does not need over elaboration.

FIGURE 2. Taxonomy for sci-tech policy events.

Before elaboration each type in detail, three points need to be called out:

- 1. This taxonomy only works for English texts. For data in other language, adaptation is required as the syntactic differences may cause other problems.
- 2. Although the 12 types are at the same level, they are still different in terms of the number of triggers and its importance in texts. The star-marked types are relatively more important than others. This will not affect trigger extraction and event classification but may be helpful for EE models in future studies.
- 3. The new set of event types is based on ACE 2005 taxonomy. As the table shows, there is a roughly corresponding relation, rather than a perfect match, between the two systems.

TABLE 3. Relations between ACE 2005 taxonomy and our sci-tech event taxonomy

ACE 2005 Taxonomy	Event Types for Science Policy Texts
movement	movement
transaction	/
life	personal life

business	economic activity
conflict	attack and conflict
	meeting
contact	opinion and talk
	publication
personnel	personnel
justice	justice
/	science and technology development
/	change in data
/	political activity

4.2.1 **Opinion & Talk.** Opinion and Talk, Publication and Meeting correspond to the Contact type in ACE 2005. Opinion and Talk refers to events that contains people's opinions, comments and remarks. In policy tracing model, those talks often have a great influence on policy making in the future, especially remarks by experts, politicians and other famous people [18-19]. The number of triggers in this type is the largest among all categories. It is also effortless to distinguish this type of event.

Both Opinion & Talk and the Contact event type in ACE 2005 contains the communication via telephone and letter, but the former also includes daily talks or verbal expression. It also has verbs that related to social media, like "tweet".

The Asia Internet Coalition, an industry group that represents Meta, Amazon, Twitter, Google and other US big tech companies, has expressed dissatisfaction with the digital competition law recommended by the Indian Parliamentary Standing Committee on Finance in December, arguing that it is "regressive" and "may dampen digital innovation in India." Indian media outlet Business Standard reported on Monday.

From <u>Global Times: US Big Tech lashes out at India; world needs coordinated</u> <u>digital rules</u>

In the above example, "expressed", "arguing" and "reported" are all verbs that express opinions. Based on Contact event in ACE 2005, Opinion & Talk contains arguments such as time, place, figure and content. For example, to describe the event triggered by "express" can be like this:

Event	Triggor			iments	
Туре	Tiggei	Figure	Time	Place	Content
Opinion & Talk	express	The Asia Internet Coalition	December	/	dissatisfaction with the digital competition law recommended by the Indian Parliamentary Standing Committee on Finance

TABLE 4. Opinion & Talk event example

4.2.2 **Meeting.** The Meeting type have something in common with the Contact type in ACE 2005 taxonomy. In ACE 2005, two subtypes are under Contact Event type: Meeting and Phone-Write, which stand for face-to-face talking and communication via phone or mails respectively. To simplify those terms, in our taxonomy Meeting type includes holding a meeting and visiting some places, while any events involving opinions and comments are labeled Opinion & Talk. Triggers for Meeting events are "attend", "host", "visit" etc. Arguments include time, place, topic, participant.

Ukrainian President Volodymyr Zelenskyy will **pay an in-person visit to** Congress on Wednesday as lawmakers prepare to **pass a budget** that includes more than \$45 billion in new military and humanitarian aid for his country.

from <u>Zelenskyy to address Congress Wednesday as lawmakers debate \$45B in</u> <u>Ukraine aid</u>

Event	т. ·	Arguments				
type	Irigger	Time	Place	Topic	Participant	
Meeting	pay a visit to	Wednesday	/	aid for his country	Ukrainian President Volodymyr Zelenskyy; Congress	

TABLE 5. Meeting event example

...

4.2.3 Science & Technology Development. This event type has no corresponding type in ACE 2005. It is solely proposed based on sci-tech policy texts. Currently this type contains the following words: discover, invent, reveal, search, calculate, indicate, examine, speculate, investigate, and so on. Phrases like "study found" are also included. It should be noted that without context, these words are less distinctive compared with verbs in Opinion & Talk event type. Their relations with science and technology development are not strong enough. However, in Science Policy corpus, they are highly related to such event and almost never appear in other event types. Thus, Science & Technology Development is one type dedicated to the current data. When data changes, it must be adjusted at the same time.

"Our findings *indicate* that compliance with WHO benchmarks compared with Australia's current sodium targets may *result in* substantial health gains and *prevent* more than three times as many deaths and new cases of disease each year," *said* Trieu. Trieu *added that* including more packaged food products and stricter sodium targets may have *had a greater impact*.

from <u>Reducing sodium in packaged foods could reduce disease and save lives</u>

This paragraph is taken from **Popular Science website**. It contains two event types:

- a) Science & Technology Development, triggered by "indicate", "have a greater impact"
- b) Opinion & Talk, triggered by "said", "added that"

We tentatively set arguments of Science & Technology Development type as: time, research or subject, finding or result. "Research or subject" indicates that the subjective of the sentence can be the researcher him/herself or the name of the research. Here we use the first sentence of this paragraph as an example:

			Arguments Finding or result			
Event type	Trigger	Research or subject	Finding or result	time		
Science & Technology Development	indicate	findings	compliance with WHO benchmarks compared with Australia's current sodium targets may result in substantial health gains and prevent more than three times as many deaths and new cases of disease each year	/		

 TABLE 6. Science & Technology Development event example

4.2.4 **Policy-making.** Policy-making event usually connects to the activity of deciding and issuing new policies. Typical triggers include "publish", "publicize", "issue", "authorize", "sign" and so forth. These triggers are featured by policy-making domain and easy to recognize. Arguments of this event type includes time, place, content, and subject.

In the Federal Register of August 12, 2019 (84 FR 39785), we **announced** that we **filed** a food additive petition (FAP 9A4823) **submitted** on behalf of Kellogg by Hogan Lovells US LLP, Columbia Square, 555 Thirteenth St. NW, Washington, DC 20004.

From <u>Food Additives Permitted for Direct Addition to Food for Human</u> <u>Consumption; Vitamin D3</u>

The Secretary of State gives notice of an Order made under Section 247 of the above Act entitled "The Stopping up of Highways (East) (No.29) Order 2022" authorising the stopping up of two areas of highway at Castleford Close, at Borehamwood in the Borough of Hertsmere.

•••

From <u>Town and Country Planning - DEPARTMENT FOR TRANSPORT</u> <u>TOWN AND COUNTRY PLANNING ACT 1990</u>

These two examples are from <u>Federal Register</u> and <u>The Gazette</u> respectively. The marked "file", "gives notice of" and "authorising" are typical triggers for Policy-making events. To describe such event, we use the trigger "filed" as an example:

Event trigger – type	tuiggou			Argumer	Arguments		
	time	place	subject	content			
Policy- making	file	August 12, 2019	/	Federal Register	a food additive petition (FAP 9A4823)		

TABLE 7. Policy-making event exam	ple
-----------------------------------	-----

4.2.5 **Justice.** A Justice event is featured by litigation and judgement in our corpus. It is roughly corresponding to the Justice Event type in ACE 2005, which has a more detailed subtypes and complicated argument slots for each subtype. Since Justice event only occupies a small amount in the corpus, there is no need to further classify it. Based on ACE 2005 arguments setting, we use a more general template to describe a Justice event, which include time, place, crime, prosecutor, defendant, other people. Here is a piece of news taken from <u>GeekWire</u>, an American technology news website:

The U.S. Department of Justice will **prosecute** the former CEO of Vancouver, Wash.-based biotech company CytoDyn, Nader Pourhassan, and an associate for allegedly defrauding investors.

Pourhassan and Kazem Kazempour, CEO of a company that led CytoDyn's clinical trials and served as its agent with the U.S. Food and Drug Administration, were **indicted** by a federal grand jury in Maryland. The indictment was unsealed Tuesday.

The pair allegedly lied about the development and regulatory status of an investigational drug, leronlimab, in order to inflate CytoDyn's stock value,

according to a DOJ statement.

from Former CytoDyn CEO and clinical trial associate indicted for defrauding investors

Table 8 and 9 describe the two Justice Event triggered by "prosecute" and "indicted".

Event type	Trigge	Arguments					
	r	Time	Place	Crime	Prosecutor	Defendant	Other people
Justice	prose-c ute	/	/	defraud	The U.S. Department of Justice	the former CEO of Vancouver, Washbased biotech company CytoDyn, Nader Pourhassan, and an associate	/

TABLE 9. Justice event example triggered by "indicted"	

Examt		Arguments							
Event type	Trigger	Time	Place	Crime	Prosecutor	Defendant	Other people		
Justice	indict	/	Maryland	/	a federal grand jury	Pourhassan and Kazem Kazempour	/		

4.2.6 **Personnel.** The Personnel event here are almost the same with the Personnel event in ACE 2005. Typical triggers include "step down", "step into", "serve on the boards", "appoint", "work as", "elect", "oust" etc., which are easy to distinguish. Arguments of this event type includes time, place, organization or employer, person, original position, current position. The following excerpt is also from GeekWire, talking about personnel movement in a tech company.

Kelly Breslin Wright is *leaving* her day-to-day operating role as president and chief operating officer at revenue intelligence platform company Gong.

Nordstrom appointed Atticus Tysen, a longtime leader at financial giant Intuit, to its board of directors. Tysen joined Intuit more than two decades ago and is currently senior vice president of product development, chief information security officer, and fraud prevention officer.

from <u>Tech Moves: Gong president steps down; Intuit exec joins Nordstrom</u> <u>board; Common Room adds product leader</u>

	Trigger	Arguments					
Event type		time	place	Organiza-ti on or employer	person	Original position	Current position
Personnel	appoint	/	/	Nordstrom	Atticus Tysen	leader at financial giant Intuit	board of directors

TABLE 10. Personnel event example

4.2.7 **Political Activity.** A Political Activity event refers to important behaviors of governments or politicians, such as establishment of institutions, diplomatic cooperation, and so forth. These are not directly linked to sci-tech policy, but still involved in policy making stage. "tally", "vote", "propagate" are all representative triggers. "Establish", "align with" are also triggers in this type, which only suggest institution establishment and alignment with other countries in our collected data.

In the country's 150-seat Parliament, 78 lawmakers, two more than the 76 needed, **voted** to oust the three-party minority government of Prime Minister Eduard Heger. Three coalition lawmakers **voted** against the government.

from <u>Slovak government falls after losing no-confidence vote</u>

The above example is about the failure of Slovak government in vote triggered by the verb "voted". The arguments are time, place, topic and participant.

Event	triggor	Argument					
type time		place	place topic				
Political Activity	vote	/	the country's 150-seat Parliament	oust the three-party minority government of Prime Minister Eduard Heger	78 lawmakers		

TABLE 11. Political Activity event example

4.2.8 **Movement.** The Movement and the following Change of Data event types together correspond to the Movement Event type in ACE 2005. (For the sci-tech policy texts, Change of Data type seems more important than Movement type.) As the name suggests, Movement type denotes the change of physical places of objects. Common triggers in this type includes "travel", "leap", "circumvent", etc. The following example is about the virologist David Evans went to Geneva for a meeting.

In November 2016, virologist David Evans traveled to Geneva for a meeting of a World Health Organization committee on smallpox research. The deadly virus had been declared eradicated 36 years earlier; the only known live samples of smallpox were in the custody of the United States and Russian governments.

from Experts debate the risks of made-to-order DNA

Arguments of Movement event type are time, starting point, destination, participant, transport/vehicle. With these the above event triggered by travel can be depicted as:

Fvent	Trigger					
type		Time	Starting point	Destination	Partici-p ant	transport/ vehicle
Movement	travel	November 2016	/	Geneva	virologist David Evans	/

 TABLE 12. Movement event example

4.2.9 **Change of Data.** It seems distinctive among all 12 types of events because data can be involved in almost any topics, but Change of Data is a crucial type in the taxonomy for sci-tech policy field. Since policy-making cannot live without status quo, events related to society and economic development are also occurred in our corpus. These events can be described in two ways. On the one hand, Personal Life, Economic Behavior and Attack & Conflict are contents of these events; on the other hand, Change of Data or data fluctuation is the statistical representation. Therefore, Change of Data is separated from other events as a unique type that does not exist in ACE 2005 taxonomy. Typical triggers include "increase", "decrease", "fluctuate", "redouble", "skyrocket", etc.

The share of US ad revenues held by Facebook's parent Meta and Google owner Alphabet is projected to fall by 2.5 percentage points to 48.4 percent

this year, the first time the two groups will not hold a majority share of the market since 2014, according to research group Insider Intelligence.

This will mark the fifth consecutive annual decline for the duopoly, whose share of the market has fallen from a peak of 54.7 percent in 2017 and is forecast to decline to 43.9 percent by 2024. Worldwide, Meta and Alphabet's share declined 1 percentage point to 49.5 percent this year.

from Meta and Alphabet lose dominance over US digital ads market

To represent Change of Data event, the following arguments are used: time, index, original number, current number and range. Take the event in last sentence as an example:

Evont		Arguments						
type	Trigger	time	index	Original number	Current number	range		
Change of Data	decline	This year	Meta and Alphabet's share	/	49.5%	1%		

TABLE 13. Change of Data er	vent example
-----------------------------	--------------

4.2.10 Attack & Conflict. The Attack & Conflict event corresponds to the Conflict type in ACE 2005, which contains two subtypes: Attack and Demonstrate. Although these events are rare in science policy, they are still inevitable in policy-making stage. For example, the following excerpt is about the vandalization of old rock art in Australian, which probably gives an alert to archeologists and policy makers to regulate human activities. Attack & Conflict is not the most important event type, yet its triggers are easy to identify, like "attack", "assault", "bombard", "destroy", "lash out" and so forth. Arguments often include time, place, participant, target, tool.

Australian authorities say vandals have destroyed rock art believed to be some 30,000 years old.

The vandals appeared to have removed parts of a barbed wire fence at Koonalda Cave and got underneath, before using their fingers to draw over the top of the Indigenous artwork, said Kyam Maher, the Aboriginal Affairs Minister of South Australia state.

...

from Australian vandals destroy 30,000-year-old rock art

TABLE 14. Attack & Conflict event example

Event	Trigger	Arguments					
type		Time	Place	Participant	Target	Tool	
Attack & Conflict	destroy	/	Australian	vandals	rock art believed to be some 30,000 years old	/	

4.2.11 **Economic Activities.** Compared with the Business event type in ACE 2005 taxonomy, the Economic Activities here covers more topics. The former emphasis the start, merge and bankruptcy of companies, while the latter also contains events that related to macroeconomics, such as macroeconomic regulation and control, funding allocations, etc. Based on the description of Business type in ACE 2005, here Economic Activities contains these arguments: time, place, participant, content.

NASA says it'll **distribute** up to \$2.45 million to 14 teams in support of experimental projects that would be right at home in the pages of a science-fiction novel — including a plan to send a flying boat to study the smoggy atmosphere and hydrocarbon-rich lakes of Titan.

•••

from <u>NASA provides funding for way-out space proposals, including a flying</u> <u>boat for Titan</u>

Event	Tuisaaa	Arguments			
type	pe		place	participant	content
Economic Activities	distribute	/	/	NASA	\$2.45 million to 14 teams in support of experimental projects

TABLE 15. Economic A	Activity event example
----------------------	------------------------

4.2.12 **Personal Life.** Corresponding to the Life event type in ACE 2005, Personal Life in our taxonomy will not be further categorize due to the limited examples in the corpus. Its triggers are featured by scientific and social development, like the verb "vaccinate". Arguments for this event type are time, place, participant.

While China counts 90% of its population vaccinated, only around 60% have received a booster. Older people are especially likely to have not had a booster vaccine. Over 9 million people older than 80 have not had the third vaccine, according to China's official Xinhua News Agency.

TABLE 16. Personal Life event example							
Event type	Trigger –	Arguments					
Event type	mgger	time	place	participant			
Personal Life	vaccinate	/	China	90% of its population			

from 'A very hard road ahead' for China as COVID-19 cases spiral

5. Conclusion. By comparing the choice of triggers in Chinese and English texts, this study provides some suggestions for handling the complex syntactic issues encountered when choosing Chinese event triggers. Moreover, with a detailed analysis of sci-tech policy texts in English, we propose a new set of event types for this closed domain based on ACE 2005 taxonomy. Since few researchers have dedicated to the standard and rationale of choice of triggers, the quality of current event extraction data is not high enough, and the performance of EE model cannot be guaranteed and evaluated. This study addresses the problem of triggers in closed domain research. Starting from triggers and texts, it works as a reference for other event extraction and classification analysis. However, this preliminary research is just a starting point in processing texts in information extraction, more studies targeting closed domain data are needed to summarize a reliable and consistent way in trigger choice and event classification, thus promoting the research of event extraction and classification.

REFERENCES

- R. Yan, X. Jiang, W. Wang, D. Dang and Y. Su, Materials information extraction via automatically generated corpus. *Scientific Data*, vol.9, no.1, <u>https://doi.org/10.1038/s41597-022-01492-2</u>, 2022.
- [2] F. Solihin, I. Budi, R. F. Aji and E. Makarim, Advancement of information extraction use in legal documents, *International Review of Law, Computers & Amp; Technology*, vol.35, no.3, pp.322–351, 2021.
- [3] Linguistic Data Consortium, ACE (Automatic Content Extraction) English Annotation Guidelines for Events Version 5.4.3. Retrieved February 17, 2023, from <u>https://www.ldc.upenn.edu/sites/www.ldc.upenn.edu/files/english-events-guidelines</u> <u>-v5.4.3.pdf</u>, 2005.
- [4] W. Xiang and B. Wang, A Survey of Event Extraction from Text, *IEEE Access*, vol.7, pp.173111–173137. <u>https://doi.org/10.1109/access.2019.2956831</u>, 2019
- [5] J. Zhao and H. Yang, Trigger-free Event Detection via Derangement Reading Comprehension, Cornell University – ArXiv, https://doi.org/10.48550/arxiv.2208.09659, 2022.
- [6] Y. Yan, Z. Liu, F. Gao and J. Gu, Type Hierarchy Enhanced Event Detection without Triggers, *Applied Sciences*, vol.13, no.4, <u>https://doi.org/10.3390/app13042296</u>, 2023.

- [7] Y. Li, C. Li, W. Xu and J. Li, Prior Knowledge Integrated with Self-attention for Event Detection, *Lecture Notes in Computer Science*, <u>https://doi.org/10.1007/978-3-030-01012-6_21</u>, 2018.
- [8] P. Li, Q. Zhu, H. Diao and G. Zhou, Joint Modeling of Trigger Identification and Event Type Determination in Chinese Event Extraction, *International Conference* on Computational Linguistics, pp.1635–1652, http://dblp.uni-trier.de/db/conf/coling/coling2012.html#LiZDZ12, 2012.
- [9] J. Liu, Y. Chen, K. Liu, W. Bi and X. Liu, Event Extraction as Machine Reading Comprehension, *Empirical Methods in Natural Language Processing*, <u>https://doi.org/10.18653/v1/2020.emnlp-main.128</u>, 2020.
- [10] D. Wadden, U. Wennberg, Y. Luan and H. Hajishirzi, Entity, Relation, and Event Extraction with Contextualized Span Representations, *Empirical Methods in Natural Language Processing*, <u>https://doi.org/10.18653/v1/d19-1585</u>, 2019.
- [11]H. Kilicoglu and S. Bergler, Effective Bio-event Extraction Using Trigger Words and Syntactic Dependencies, *Computational Intelligence*, https://doi.org/10.1111/j.1467-8640.2011.00401.x, 2011.
- [12]L. Tian, W. Ma and W. Zhou, Automatic Event Trigger Word Extraction in Chinese Event. *Journal of Software Engineering and Applications*, https://doi.org/10.4236/jsea.2012.512b040, 2012.
- [13]Q. Deng, S. Xie, D. Zeng, F. Zheng, C. Chen and L. Peng, An Event Extraction Method for Public Security, *Journal of Chinese Information Processing*, vol. 36, no. 19, pp.93-101, 2022.
- [14] W. Wang and X. Liu, Spatiality and temporality: The fundamental difference between Chinese and English, *Lingua*, vol.251, https://doi.org/10.1016/j.lingua.2020.103011, 2021.
- [15]N. Chen, and X. Li, An Event Extraction Method Based on Template Prompt Learning, *Data Analysis and Knowledge Discovery*, 2022.
- [16] The White House, Office of Science and Technology Policy, https://www.whitehouse.gov/ostp/.
- [17] Technology: Research & Analysis | CSIS, https://www.csis.org/topics/technology.
- [18]X. Wu, Research on the Key Technology of Automatic Generation of Tracing-oriented Evolutionary Summarization, Master's dissertation, Peking University, 2021.
- [19] W. Li, Understanding the Decision Making Approach to the Policy Process: Theoretical Foundations, Evolution Process and Future Trends, *Journal of Gansu Administration Institute*, vol.124, no.06, pp.46-67, 2017.