Cognitive Corpus Studies: A New Qualitative & Quantitative Agenda for Contrasting Languages

Barbara Lewandowska-Tomaszczyk[1]

Abstract

This paper is an attempt to find a place for contrastive studies in the present-day linguistics. It focuses on some philosophical and linguistic assumptions of Cognitive Grammar, which is relevant to studying contrasts between languages. Two of the fundamentally important concepts discussed in the paper are the concepts of *equivalence* and its philosophical 'anchoring point' *tertium comparationis*. The first part of the paper presents a debate on their definitions and interpretation, and a new, evolving perspective in terms of a cognitive corpus linguistic paradigm. Introduced here is the concept of a communicative shift in meaning, or *reconceptualization*, in terms of the speaker's and addressee's *approximation* to their universes of thought. What is proposed in this paper is a new look at and a research agenda for the concept of equivalence in contrasting languages, based on two sets of criteria. The first set is qualitative and mental in nature, serving as a crucial function for the entity of *Event*, while the second is quantitative, capturing *distributional* and *frequency facts*, which help to identify the (proto) typical and increasingly peripheral semantic construal-types in the contrasted languages. The discussion is exemplified with English and Polish corpus data of participial modification.

[1] University of Lodz, Poland

Keywords: approximation / cognitive linguistics / construal / contrastive studies / equivalence / frequency of use / language corpora / participial modification / reconceptualization / tertium comparationis / translation

1. Introduction

Since the 1960s, similarities and differences across languages have been subject to close scrutiny, first in the framework of transformational-generative grammar and its further developments, and later, with the change of the paradigm towards more cognitively-oriented studies, in terms of varieties of cognitive grammar.

The perennial problems in linguistic comparisons since their early attempts have been, first, the question of a point of reference, or an anchoring entity, which would ensure the comparability between the structures juxtaposed, and, secondly, the problem of purported equivalence between them.

This paper is yet another attempt to look at language contrasting and to provide an answer to these queries through an approach that makes use of the relevant concepts of Cognitive Linguistics as put forward in the works of George Lakoff (1987) and Ronald Langacker (1987, 1991), and enriched by the ideas developed by Lewandowska-Tomaszczyk (2008, 2009). This study posits and illuminates the notion of *Event* on the one hand and the concepts of communicative and translational *reconceptualization* and *approximation* on the other.

Parallel to these linguistic qualities considered the basis for contrastive studies, recourse to quantitative properties underlying linguistic structures across different systems has been considered indispensable if a fuller account of cross-linguistic comparison is to be established. The quantitative factors comprise frequencies of occurrence of language items and distributional facts related to them from which to conclude the nature and structure of linguistic meanings.

The reference points in contrasting languages are, therefore, both qualitative and quantitative in nature (cf. Lewandowska-Tomaszczyk & Dziwirek, 2009). The starting point in such studies is identified in the form of prototypical Event Scenarios and their conceptualizations as constructed by language users in a given cultural community. Such contrasts will be discussed here on the example of participial modifiers in English and, moreover, with reference to authentic language data from two languages, English and Polish, as used in widely referenced corpora. In conclusion, an integrated research agenda for contrasting languages will be presented and described.

2. Qualitative dimensions in comparing languages: Commensurability criteria

In his seminal publication *Women, Fire and Dangerous Things*, Lakoff (1987) proposes four types of what he calls *Commensurability Criteria:*

(1) truth-conditional criteria (classical translatability);

(2) criteria of use;

(3) framing criteria; and

(4) conceptual organization criteria.

An attempt to decompress the criteria into smaller principles leads to a conclusion that the Commensurability Criteria make it possible to contrast languages according to particular frames of reference. The first criterion involves the truth-condition principle. According to this criterion, the language user should be able to provide systematic rules for computing the truth conditions of a sentence by assigning a reference and a truth value to elements of the sentence in L1 and state whether they are identical to a construction considered equivalent in L2. In other words, one has to know the conditions under which the sentence in L1 and the sentence in L2 are true and be able to state whether the sets are identical or different.

The criterion of use involves a distributional range of language elements. The extent to which such lexical elements in English as to put down and to lie down can be considered synonymous is weakened when the distributional criteria are taken into consideration. The former refers in most cases to a physical action, while the latter is predominantly used in a metaphorical sense. Furthermore, the former has a higher frequency of occurrence in more varied contexts. A similar procedure and criterion are used to contrast interlingual synonymy, i.e. so-called equivalence.

The framing criterion combines the linguistic knowledge with the knowledge of the outside world and imposes on particular language units object or event schemata, which regulate a top-down perspective on individual meanings. The difference between the FRUIT schema and the VEGETABLE schema with respect to tomatoes is the case in point.

And, finally, the conceptual organizational criterion regulates the perspective of an object within a given category. The most interesting cases

here will be cases of polysemy, which may be differently organized in one language than in another, due to a distinct conceptual organization of the relevant senses they comprise (e.g. the polysemous nature of the form *chest* or *bright* in English is not necessarily present in their equivalents in other languages; cf. Lewandowska-Tomaszczyk, 2007). Numerous instances of conceptual gaps, as, for instance, in culture-specific terms (such as Christmas pudding, whose concept is absent in may cultures, or *annoyance*, which is common in English and not identically conceptualized in many other systems), can exemplify this phenomenon. So would referential descriptions of the concepts which are considered as L1 – L2 dictionary equivalents, such as the division of a 24-hour unit into the phases of day and night, and its subdivision into smaller parts like noon, afternoon, twighlight or dawn, which do not find exact correspondences in other language/culture systems.

A theoretical possibility connected with the above criteria would require that the languages in question be equal in all aspects—that is, they would need to be totally identical, or the s*ame* language. Another extreme theoretical option will be the languages which would satisfy none of the above criteria, such as the languages of the Quinean *gavagai* type, with no common platform to refer to. In reality, the language systems turn out to be partly "calibrated", which represents a typical cross-linguistic situation with what can be called *equivalents* not of an *identity* but rather of an *approximation* type.

The approximation alluded to above (cf. Lewandowska-Tomaszczyk, 2012) has its expression not only in the *content of a linguistic unit* but is also captured by its *constructional properties.* It is precisely the viewing arrangement of the scene, i.e. the *construal relations*, which appear to be

crucial in contrasting the semantics of *Events* in one or more linguistic system. The difference between the semantics conveyed by a complete sentence, e.g. *Peter swallowed the candy*, and that of a corresponding nominalization, which imposes a more reified frame on the content—*Peter's swallowing of the candy*—is a difference in one of the possible construals of a scene. Crucial to the notion of cross-linguistic comparison is also the concept of *profiling*, in which a profile of an expression is, to quote Langacker (1991, p. 551), "the entity that the expression designates, a substructure within its base that is obligatorily accessed, accorded special prominence, and functions as the focal point within the immediate scope of predication". Thus, profiling is an aspect of construal, in terms of which semantic differences can be accounted for in the same language or in the comparison with other linguistic systems.

Yet another type of a comparison involves figurative usages, i.e. mapping operations of one domain onto another domain, or part of a domain onto the whole domain, etc. Fear relations, for instance, in a fear-event can be accounted for in many cultures by resorting to the concept of *force dynamics.* Force dynamics (cf. Talmy, 1988) describes the ways entities interact in an event where one of them is trying to exert power over the other using a physical (prototypically) or a mental force. Fear is conceptualized by assuming a scenario in which fear is perceived as an agonist ("doer") and the experiencer as an antagonist ("affected"). The outcome of the force dynamics depends on the balance of forces: either the agonist wins and fear overcomes the experiencer, or the experiencer succeeds and fear is conquered (cf. Lewandowska-Tomaszczyk, 2010; Wilson, 2010). And yet, even though a large number of basic metaphors are common across languages, the figurative

frames (Source Domains) are not always identical in different cultures and languages. In the Indonesian *kepala mobil* "front (lit. 'head') part of car(s)" and *pantat mobil* "back (lit. 'buttocks') part of car(s)", the concepts of head and buttocks will be used, while in Slavic languages (see the Polish *czoło*) the concept of the *forehead* would be used in the former function, while in English the body part *back*, an extension itself, will stand for the latter. Nevertheless, the process of metaphorization will invariably be a human universal cognitive ability which can serve as a legitimate frame of reference in looking at language contrasts.

3. Quantitative parameters in language comparison

Apart from the qualitative dimensions discussed above, a second large group of parameters, *quantitative* linguistic criteria, consists of the following:

(1) frequencies: (i) in general language, (ii) in a context-specific language variety;

(2) quantitative distributional facts;

(3) sentence length;

(4) type/token ratio;

(5) lexical density (low frequency-high frequency); and

(6) naturalness (frequency and contextual preferences).

Frequencies in general language use are quantitative data usually obtained by looking at the frequency ranks in large language corpora. For instance, the infinitive *be* in English has the frequency of 581,623 occurrences in the British Corpus of English (BNC) in the 100-million-unit data, while the

corresponding French *être* is found 18,402 times in close to 14-million-unit-long French data (IntUne French corpus). Taken at the normalized values, the British data will give slightly over 58,000 occurrences for every 10 million units; in the French data, on the other hand, the infinitive will occur approximately 13,500 times in a similar corpus of 10 million samples. In other words, the French infinitive is almost 4 and a half times less frequent than its English counterpart *(to) be.* Another question to be asked concerns the reasons for such a huge difference. And here we come to the qualitative functional and distributional analysis of the verbs in question, which will illuminate the relevant areas of differences. One of the reasons is the fact that the Future Tense in English is also formed using an infinitive *be* (*I will be here at three o'clock tomorrow*), whereas the French equivalent employs a separate form for the Future Tense (*Je serai ici demain à 11 heures)*.

Quantitative distributional facts related to contextual factors can provide new insights into contrastive language studies. First of all, context disambiguates the senses; for example, the form *cream* is disambiguated into *coffee cream* and *facial cream*, or the form *rabbit* activates a different image in a *rabbit in the bush* and *rabbit in wine and garlic sauce.* However, what matters additionally is the frequency values of such cases, both in one language and cross-linguistically.

The frequency characteristics will also illuminate a qualitative factor with respect to the examined data, namely, the degree of *naturalness* associated with individual constructions. The English gerundive structure in *my having painted the house a very special shade of yellow was hard work*, juxtaposed to the semantically close *I painted the house a very special shade*

of yellow and it was hard work, tells us nothing about the *usage-based* parameter of either form. More revealing in this case is the concept of *naturalness*, understood as a system of the speaker's/writer's preferences of the use of a language unit, which is expressed via the frequency of its occurrence in a well-defined context (Lewandowska-Tomaszczyk, 2001, p. 178; Dziwirek & Lewandowska-Tomaszczyk, 2010, p. 128). A more natural unit/structure then will be the one used more frequently in a given context, i.e. the *and*-conjoined construction in the above example.

The research task involving a *cross-linguistic comparison* is thus built around identifying a similarity as *a dynamic notion*, represented as a cline exhibiting a gradual increase in diversification. *The degree of equivalence* between L1 and L2 structures can thus be measured in terms of the reference categories mentioned above such as the typology of the category of *naturalness*, as well as categorization levels, prototypicality, image-schemata and their extensions, profiling and construal relations of various types.

What expresses an asymmetry between languages is a *displacement of senses* (Lewandowska-Tomaszczyk, 1987). The displacement of senses originates from the presence of *referential, conceptual* or *lexical gaps* in one language and accounts for the semantic/syntactic mismatches, such as the absence of the superordinate category GO in the case of the Polish verbs of movement and the syntactic structures it introduces, or the absence of a lexicalised distinction between "striking with a foot" and "striking with a fist" in French, compared with the forms *kick* and *punch* in English, as illustrated below.

Eng. kick Fr. *donner un coup de pied 'strike with a foot'*

Eng. strike { Fr. *donner un coup*

Eng. punch Fr. *donner un coup de poing* 'strike with a fist'

The approach to contrastive analysis advocated in this study is usage-based. Each verbal event can be described by means of a conventionalized set of schemata (i.e. a set of common properties abstracted from a number of such events), characteristic of this particular act (cf. Lewandowska-Tomaszczyk, 1987). Meanings of individual lexical items involve various types of (lexical and/or technical) *instructions of discourse incrementation* (cf. Seuren, 1985), which direct the items to their positions in discourse or outside the discourse domain.

Looking at an individual lexical item from the perspective of a system, one can identify its meaning in terms of its *multidimensional networks of meanings*, which reflect its distributional characteristics and position in the system, e.g. synonymy and oppositeness, or polysemic links. From the usage perspective, some of these dimensions are more salient than others. Discourse is an active factor in meaning construction. It can reinforce some and weaken other dimensions. By employing such contrasts, what is obtained in context is a higher monosemy of the sign. The multidimensional entities, which express linguistic meanings, are only partially equivalent in different languages. They uniquely activate further dimensions, not necessarily overlapping in different linguistic systems, e.g. the polysemy of the English *chair*, ranging from "a piece of furniture/a seat"; "an official position"; "a person holding such a position"; to

"professorship", as opposed its counterparts in many languages, in which *chair* (e.g. Pol. *krzesło*) refers predominantly to "a piece of furniture/a seat".

Language units, words, phrases, and sentences repeatedly used in discourses are eventually abstracted from their use and considered conventional constructions (cf. Langacker, 1988). The framework we have been employing in our work is both cognition- and construction-based in this sense. All the data come from authentic language use, corpus materials in both languages. The interpretation is cognitive as it assumes conceptualization principles, which underlie linguistic activities, and interactional to the extent that meanings in languages immersed in large knowledge and culture frames, together with discourse context, can be overlapping, but never identical.

Different degrees of contrastive correspondences in the languages also represent what can be conjectured to be "approximations", as I call them (Lewandowska-Tomaszczyk in press), or in some acute cases "mismatches" (cf. Dziwirek & Lewandowska-Tomaszczyk, 2010). They signal not only the differences in cultural and discursive contextualization in the systems contrasted, but also express and symbolize the language-specific senses and individual variation.

4. Universal tertia comparationis

The search for the properties which would anchor down a cross-linguistic comparison is curbed by the fact that there is less to be found in the world languages that could be considered substantially identical. Rather, what is observed is a contrastive skeleton, or frame, in which certain properties are a constant. What can be predominantly identified are *cognitive tertia* on the one

hand and *universal procedural* and *structural universals* of different types on the other.

4.1 Cognitive tertia

Cognitive Tertia Comparationis in comparing languages cover a number of human cognitive abilities and involve analogy, abstraction, metaphorization, as well as combinatorial powers, or what can be dubbed Chomsky's *recursion* properties (cf. Hauser, Chomsky, & Fitch, 2002).

The basic cognitive parameter subsumed under the human capacity of analogy and abstraction belongs to the ability of *categorizing* objects and phenomena and its main attributes, such as the representation in terms of basic image schemas, schematic category structures, comprising prototypical and peripheral category members, combined into larger Idealised Cognitive Models, culturally and contextually bound (Lakoff, 1987). The criterial feature of these structures is their partial compositionality and the presence of on-line meaning building mechanisms in terms of *emerging structures* (Lewandowska-Tomaszczyk, 2010).

I propose that the *universal processes* in cross-linguistic tasks such as cross linguistic comparison, transcultural communication, or translation, are comprised of two inseparable elements. Firstly, they involve *reconceptualization of the incoming material*, which invariably leads to the second element, *conceptual approximation* of the output material. These elements function both with respect to the outside world, as no representation, be it linguistic or non-linguistic, would cover *all* parameters and details of the reality, and with reference to *L1* in relation to *L2* (where *L* designates any language or linguistic variety used by Speaker 1 to communicate with Speaker 2, and in reply,

Speaker 2 addresses Speaker 1). In other words, no linguistic or any other semiotic representation will be the *only full mirror of the outside world*. A linguistic structure is an outcome of a number of cognitive operations starting with the parameters of construal, focusing, perspectivizing, etc. (cf. Langacker 1987, 1991).

Cultural impact, where *culture* is understood as *conventional* (i.e. *shared imagery and practices*), cannot be ignored (cf. Lewandowska-Tomaszczyk & Wilson, 2011). Neither can we ignore linguistic typological frames of reference, which give rise to language-specific *constructional* and *semantic frames* with a range of distinct analysability criteria and construal principles, including degrees of prominence of a scene, action parameters, figure/ground relations, degrees of schematicity (cf. the coarse- vs. fine-grained picture), scope of predication, and force-dynamic relations. Typologically distinct linguistic construal types are an outcome of a comparison of two or more linguistic systems.

Last but not least in the present inventory are *pragmatic* and *interactional effects*, both of which are part of contrastive discourse analysis, whose outcomes complete the picture of a cross-linguistic analysis of two or more systems.

4.2 Reconceptualization cycles, approximation and tolerance spaces

Communication involves a number of *cycles* of reconceptualization of an original message (Lewandowska-Tomaszczyk, 2010). These reconceptualization cycles lead the communicator to set up an *approximative potrayal* of a compared scene within a certain *tolerance space*. Such a linguistic phenomenon as polysemy, as well as what I generally refer to as *a cubist*

portrayal of the outside world in language, supports the claim that meanings must be networked within a certain *tolerance space.* In communication, in translation, and in contrasting languages, tolerance spaces are dynamically construed up to a (context-specific) *tolerance threshold*, beyond which a miscommunication or a communicative boycott occurs in actual communication and in contrastive studies, and beyond which *no* cross-linguistic *similarity* or *resemblance* can be posited. As also proposed by Peter Gardenförs (2004), semantic representation of a concept within a given *conceptual space*—which may be understood, as a set of quality dimensions, or separable, as in *shape*, or integral, as in *colour* and *shape*—is curbed by a set of certain constraints on sense divergence, that is, their tolerance thresholds which represent the boundaries in communicative interactions and which are limited by the intra- or inter-systemic variety of particular linguistic signs used in communication. The tolerance measures imply resemblance, which is either conventional culture- and context-specific, or else unique to a given speaker.

5. Event

What I want to propose in this paper is that one of the few substantive tertia comparationis in cross-linguistic comparisons be posited in terms of the mental entity of *EVENT.* The first question to be asked, namely whether events constitute a coherent metaphysical category (Lewandowska-Tomaszczyk, 2011), cannot be unambiguously answered. However, a number of dispositions, which can be considered regulative for the ontological category of events (cf. Zacks et al., 2001), can still be identified. They include perceptual criteria for infants' perception, discrimination and counting of events, action-

based observations in animates' planning and executing of actions in contexts, linguistic factors related to the devices dedicated to describing events, as well as the mental layer—as thinking about many aspects of the world, its properties and actants in terms of places, time, causes, consequences, etc. requires framing in the form of an event structure.

Classification of events into different types can be considered a structural skeleton of cross-linguistic descriptions. Vendler's first typology (1957) into *activities* (defined as *a homogenous event with no natural finishing*), *accomplishments* (defined as *a non- homogenous event with a culmination*), *achievements* (defined as *a culminating event*), and *states* (defined as *a homogenous event which may extend over time*) is only the beginning of a debate on this issue. Von Wright (1965) considered it necessary for an event to have a begin-point and end-point, and everything that happens between a negative value of a unit [-φ] and its positive counterpart [φ] with a transition operator underlying the logic of change. Donald Davidson, particularly in his seminal work on action and events (1960), proposed obligatory criteria of spatiotemporal unity and causality for events in terms of his linguistic semantics of action. Davidson's approach has been reigning supreme for many decades now and has given rise to a number of subtle typologies of events. A more holistic picture can treat event as subsumed in terms of chains of subevents, which start from a *Stative phase* > change into *Inchoative* > *Processual* > *Terminative* > and transform into *Stative* again. In other words, it represents a *Change of phases*, where either the whole event or any of the fragments can be conceptualized in different language systems and by communities of language users.

5.1 Conceptualization of events

Events understood as *phenomena that happen* can be prototypically conceptualized as one unfolding and gradual entity having its beginning and ending. More heterogeneous events can be perceived as a gradual durative s*equence of sub-events*. However, an important question whether sequential subevents, for instance, in Kalam, a language of the highlands of the Papua New Guinea, or Thai serial verbs, for that matter, constitute separate conceptual entities subsumed as one *meta-event* or the series of subcomponents are only structural and have no impact on the holistic or elemental perception of an event in question, is still not resolved and requires further research (Givón, 1990; Pawley, 2011). There are, however, other conceptualization frames possible for events as well. Events can be conceived as *things*, and hence can be reified to different degrees and eventually perceived also as an *attribut*e in different phases of inception, duration, completion or iteration (Lewandowska-Tomaszczyk, 2011.

Events can thus be perceived similarly to objects; they are patronymic—having their parts—but, unlike objects, they have a temporal dimension. Even when forced into object frames (as gerunds are in English), they can retain their aspectual properties. They can also be attributivized to a different extent (as participial modification in English) with the degree of their adjective-like properties more or less transparent. In English, the attributive, stative properties of (verbal) participial senses are marked with their fully adjectival prenominal position, or, when used postnominally, as a more occasional property, participial modifiers frequently indicate contemporality with the time of an utterance. Languages differ in the ways of conceptualizing

events and ways in which they segment the world of events into smaller units as well as in the accessibility of the morphological and syntactic resources to signal these cognitive operations.

6. Prototypical events & asymmetric events

A prototypical event is usually an action performed by an agent. The question of whether intransitive events with no objects present or transitive ones are more prototypical remains to be seen. Langacker (1991) opts for the prototypicality of an action in which an agent is using some force to act on an object, i.e. clearly a transitive action with a display of a full causality scenario, which is for some the gist of an event characterization. However, if accessibility of different structure types in language evolution could be considered part of the prototypicality criteria, it is intransitive, no-direct object addressed action such as movements that can rather be considered to be the primary, evolutionally basic frame of reference, at least in some language evolutionary models (e.g. Provogac, 2010)[2].

A Transitive Action Event is an event which portrays a transmission of force, or energy, among the event participants (cf. Talmy, 1985; Langacker, 1991) and involves a temporal dimension then. The same event can be perceived as a series of subevents, frequently involving a more homogeneous sequence, such as *I read a book*. Processes, however, represent a less prototypical type (e.g. *withering*)—they lack bounding of different sorts, even

[2] Compare: "Transitive clauses involve additional layers of structure, and can be hypothesized to have been a later evolutionary innovation". (p. 238).

though they can also lead to a change of state, similarly to temporally bounded events as in the verb *break*. An interesting point is that, both before and after the bounding, what can be presupposed is the presence of *states*, which can curb a process and transform it into an event. Events then bear in themselves a potential for any type of activity, action, or eventuality, as some refer to them, not necessarily a prototypical form of transitive or intransitive action chain. Taken from this perspective, *State*s can be considered contextual frames of events while *Accomplishements* and *Achievements* refer to the final phase of events. Properties which can be designated in an event structure can be referred primarily to a type of event as well as to its part profiled in the linguistic unit, which expresses a given phase. Furthermore, the type of an action involved as well as its temporal frame, expressed cross-linguistically by a variety of markers, are the criterial factors in the analysis. As English does not have the grammatical tools to mark all the phases (e.g. inchoative), so the schema below only illustrates approximate structures in English.

Stative phase

Change

Change of state (punctual) *break*

Gradual

Homogenous (*grow*)

Heterogenous (*eat*)

Frequencies:

Single acts [bounded] *He kicked the ball*

Frequentative [bounded] *Tom used to visit us*

Durative [unbounded] *He is reading*

Construal types:

Sequential scanning *Olivia is walking to the garden*

Summary scanning

Substantive (nominal, gerund) *her walk*, *her walking*

Attributive (adjectival) *reading students*

Attributized gerund *a walking stick*

Event phases:

Inchoative (inceptive) phase *launch*, *going pale*

Durative phase *Mark is riding his bike*

Terminative phase (telic or atelic) *Jeremy pushed the door open*; *Peter read an email*

6.1 Symmetry and asymmetry in perception and expression

Events, in perception and linguistic expression, can be treated as symmetrical entities, when two or more events or their parts are perceived as two or more parallel units or appear in a sequential order *(symmetric* events). Alternatively, they can be perceived and linguistically expressed as what I call *asymmetric* events (Lewandowska-Tomaszczyk, 2008). Asymmetric events can be considered a substantial universal and cover the material referring to two or more events of unequal statuses in an utterance, for example, the forms used in sentential complementation and nominalization, in relative, adverbial, and modifying constructions, or, in some languages, in *semantic asymmetries* in what looks like fully balanced coordinate constructions. The idea behind the asymmetry is that in different world languages, such system differences occur between fully elaborated events and those which are desententialized and lose or lack their assertive force (Cristofaro, 2008). Languages of the world display

different construals of the asymmetry and involve various morpho-syntactic means to express it even though the concept of asymmetry seems to be present in all of them.

The asymmetries, then, can involve a perceptual (and linguistic) transformation of a sequential (process or action) into a construct reified to different degrees, i.e. a thing, which can be then perspectivized as an attribute of varying degrees of strength expressed in terms of participial and adjectival constructions, as in the majority of European languages. However, some Asian languages, such as Chinese, have no participles or participle-like constructions in their systems. But even in such languages some other (lexical or contextual) markers are used to substantiate asymmetry functions. It is usually adverbial phrases added or contextualizing information that can generate an interpretation similar to the English participial constructions*. The aspectual system in Thai, on the other hand, is quite complex—the durative and progressive aspects involve two aspect markers, which can co-occur in some contexts and are not used in some others. For some phrases which would involve postnominal modification in English, such as *a man drinking*, the structure used in Thai is that of a noun followed by a relative clause with the aspect marker *kamla*ŋ. However, for the English lexicalized passive participle *drunk*, as in *a drunk man*, Thai either uses a different lexical form *maw* or, if the word for *drink*, i.e., dɯ̀ɯm is employed, followed by the *completive aspect* marker, lɛ́ɛw, which can function as an equivalent of the English *already*. The phrase dɯ̀ɯm lɛ́ɛw then means literally *has already drunk*, and not *be drunk* as in English**. On the other hand, a comparison between English and Polish, one of the Slavic languages in Central/Eastern Europe, with reference to similar

asymmetries characterized by a different construal, will be presented in the section to follow.

The grounding of asymmetry markers understood as categorizing temporal, spatial, etc. dimensions can be either weak, as in the English verbal noun (*proposal*, *construal*) or more transparent or stronger, as in the English gerund *John's having submitted his thesis too late*, in which the perfective completion of the act is syntactically marked. When put in a finite sentence e.g. *John's having submitted his thesis too late is a problem* the desententialized construction displays its asymmetric status, vis-a-vis the fully finite sentential part *is a problem.*

7. A sample of English and Polish contrasts

Our short analysis of the similarities and contrasts between English and Polish will be exemplified by reference to present participles, gerunds and participial modifiers. The English samples have been acquired from the 100-million-word British National Corpus and a smaller 15-million Longman and Microconcord Sampler for English. The Polish samples have been obtained from the National Corpus of Polish (www.nkjp.pl), which covers over one and a half billion segments at present, and from two smaller (10-million and 20-million) PELCRA corpora. In other works on English-Polish contrasts (Lewandowska-Tomaszczyk in press), I also resort to *bilingual parallel (translation) corpora* as another important source of cross-linguistic data.

7.1 Present participle

The ambiguous sentence (1) in English has to be disambiguated in Polish as (1a), while the English (1b) would require the change of the modifier position from the prenominal to postnominal one in Polish.

(1) A boy looked at a girl **reading** a book
[noun complement present participle
(present participial clause)] [a. *I-I*, b. *I-Mary*]

(2)

(a) Chłopiec spojrzał na dziewczynę **czytającą**
boy looked at girl reading
[**adv**, coreferential with main subject] książkę – adverbial participle

(b) Chłopiec spojrzał na ?**czytającą dziewczynę książkę**
boy looked at reading girl book
/dziewczynę czytającą książkę
girl reading book
[**adj** Acc Sg], non-coreferential with main subject]
attributive participle

Co-temporal prenominal modification in the form of a participial/attributive (or some cases gerundive) construct in English (3) has a symmetric parallel in Polish (4). However, if an object of the construct is used as in (5) the sentence will be ungrammatical in English but remains grammatical in Polish (6). The construal of the English and Polish events in (3) and (4) is more attributive than

in (5) and (6), with a more sequential verbal scanning profile, not used in English but a regular formation in Polish.

(3) This was a frightening scream

(4) To był przerażający krzyk (more attibutive/property)

(5) This was a frightening * us all scream

(6) To był przerażający nas wszystkich krzyk (more verbal/sequential)

Presented below are the frequency data of a number of conceptual profiles of the English modifying form *drinking* identified in the English samplers.

(7) *drinking* Conceptual Profiles[3]

[15million units/569 occurrences/161 occurrences in modifying functions]

(i) Pre-modification (69)

(a) Non-co-temporal/summary/attributive/habitual (6):

He isn't a drinking man

Indeed their condemnations of drinking mothers are particularly sharp

(b) Summary/reification/stative (63)

The size of a drinking straw

I shrug my shoulder and walk to the drinking fountain

[3] Numbers given in brackets indicate the frequency of occurrence of the form in the English or Polish corpus samplers.

(ii) Postmodification (91)

(a) Co-temporal, processual [single] (61)

Jeweled women, drinking Turkish coffee

(b) Non-co-temporal (generalized), processual/ repetitive/bounded (22)

Only sitting up at night and forever drinking

(c) Unbounded (2) /repetitive construction

when I used to be drinking good ale.

(d) Modified attribute (6)

heavily drinking

(iii) Causative (3)

(a) Cognates

pit-dirt, dinnerless, some mile away from home, across the darkness, drinking himself drunk Paul stood in the doorway.

(b) metaphoric *into*

is the unfaithful, deceitful {friend} who leads Leonardo astray into drinking, gambling and having romantic affairs.

With the use of the WordSmith Tools, patterns involving the form *drinking* can be generated from the English sampler concordances (8) as well as relevant clusters (9), which indicate the most frequent objects of drinking in the English corpus texts in the descending order: *water*, *coffee*, *beer*, *tea*, *wine*, *champagne*, *whisky*, as well as co-temporal activities performed such as *eating* and *smoking*.

(8) Patterns of *drinking*

N	L2	L1	CentreR1	R2
1	AND	DRINKING		AND
2	EATING			BEEN
3	HEWAS			A
4	OF OF			IN
5	HAD	THE		IT
6	AND	FOR		WATER
7	I WERE	WITH		
8	TO IS			HE
9	S S			FROM
10	A HIS			TEA
11	YOU	COFFEE		OUT
12	WE	WITH	I	BEER
13	IN A			TEA
14	THEY	HEAVY		SO
15	WAS	STOP	HIS	WINE
16	SMOKING		OUT	BUT
17	HIS	IN		SMOKING
18	HIM	SAT		CHAMPAGNE
19	BY WITHOUT			WHISKY
20	SHE	FROM WINE		

(9) Clusters *drinking*

N	Cluster	Freq.	Length
1	EATING AND DRINKING	21	3
2	HAD BEEN DRINKING	13	3
3	HE WAS DRINKING	12	3
4	DRINKING IN THE	10	3
5	HE S DRINKING	8	3
6	HE HAD BEEN	6	3
7	THE DRINKING OF	6	3
8	WE WERE DRINKING	6	3
9	OF EATING AND	5	3

N	Cluster	Freq.	Length
10	OF HIS DRINKING	3	3
11	SHE WAS DRINKING	5	3
12	OF DRINKING AND	5	3
13	DRINKING FROM THE	5	3
14	DRINKING TOO MUCH	5	3
15	DRINKING AND SMOKING	5	3
16	I VE BEEN	5	3
17	AND DRINKsING AND	5	3

For contrastive purposes, the data on the passive participles are also presented. The passive participle *drunk* has two basic grammatical functions. Firstly, it is used as a part of the Perfect aspect of the verb (*he has drunk three glasses of beer*) and secondly, it is part of the more lexicalized passive construction in the sense of *excessive drinking*. The table below (10) presents the *drunk* clusters, where the more lexicalized sense is clearly prevailing. This observation seems confirmed in the data in table (9), where the clusters of the form *drinking* show a more frequent (lexicalized) gerundive variant than a corresponding participial form.

(10) *Drunk* Clusters

N	Cluster	Freq.	Length
1	HE WAS DRUNK	22	3
2	TOO DRUNK TO	16	3
3	HE HAD DRUNK	11	3
4	DRUNK IN THE	11	3
5	TO BE DRUNK	11	3
6	I WAS DRUNK	10	3
7	GOT DRUNK AND	10	3
8	A LITTLE DRUNK	8	3
9	TO GET DRUNK	8	3

N	Cluster	Freq.	Length
10	DRUNK AND I	7	3
11	WAS DRUNK AND	7	3
12	YOU RE DRUNK	7	3
13	DRUNK TO REMEMBER	6	3
14	ARE YOU DRUNK	6	3
15	AS DRUNK AS	6	3
16	WHEN HE WAS	6	3

The patterns generated from the concordances can give the researcher additional information, concerning the (direct) objects used with the Verb *drink* in English and the corresponding *piĆ* in Polish (11). A contrastive task is to compare (11) with a similar table of patterns generated for English (8).

(11) *piĆ* patterns in Polish [translated into English are content words]

N	L2	L1	Centre	R1	R2	
1	NIE	JEŚĆ	PIĆ	I	I	'eat'
2	Z			NIE	SIĘ	
4	SIĘ	JEŚĆ		ALKOHOL		'eat'/'alcohol'
5	Z	ZACZĄŁ		WÓDKĘ		'begin'/
						'vodka'
6	Z	DO	TO			
8	BY	WOLNO		PIWO PALI		'beer'/'smoke'
9	CZY	MAM		ALKOHOLU		'alcohol'
10	JA	BĘDZIEMY		Z	GDY	
11	ŻEBY	BY		W	ALE	
12	WOLNO		MU	NA	TYLKO	
13	PO	MU	JU	A	ZACZĄŁEM	'begin'
14	MOGĘ		POCZ	WINO JEŚĆ		'wine'
15	TRZEBA	BĘDZIE		WODĘ		'water'
16	CHCIAŁO		CO	WODY [...]		'water'
19		PRZESTAĆ		KREW		'blood'

The frequencies of the Polish present participial *pijąc** types (12) are different from those in the corresponding data in English, as presented in (8). Polish as an inflectional language allows a more flexible word order of sentence units than English does, so Polish grammar identifies constructions in (12i) and (12ii) as equally acceptable, although the left-branching of modification turns out to be less frequent, most probably for language processing reasons. The result is that the Polish frequencies of prenominal and postnominal modification are not the same (values given in brackets):

(12) *pijąc* Polish construction patters

Source NKJP

(i) Prenominal *pijąc* NP. [4]*

pijące gromady 'drinking groups'

Nałogowo pijące wyrostki 'compulsively drinking teenagers'

(ii) Postnominal NP. *pijąc* Obj NP. [75]*

Osoby pijące alkohol 'persons drinking alcohol'

Osoby nadmiernie pijące 'persons drinking excessively'

7.2 EVENT

Event phases constitute yet another case where an event in a language-specific construal and its profiling play an important role. For instance, in English the action of *opening* can capture the following phases of opening:

(13)

(i) Cordelia ***was opening the room slowly. She opened the door*** *and went in.*

(ii) *Cordelia rummaged in the fridge and brought out* ***an opened package*** *of store doughnuts.*

(iii) *She looked* ***through an open window of the building***

The Event Structure presented in (13) involves the act (process) of opening and its terminative phase in (i), the resultative phase with the negative presupposition in (ii) and the final state in (iii). The frequencies of the particular uses are as follows:

(14)

EVENT of opening [literal and metaphorical]

BNC –100 mln

Y closed [initial state] // X is opening Y [Y opening; opening Y] → X opened Y [event] → [28,562]

Y is opened [terminative phase] →

Y opened [90, (Mod)]→ opened Y [120, (Mod)] [terminative attribute] →

open Y [final state] – [10,822], (NP. open [1,700], open NP. [1,900])

The terminative phase can also be marked in a causative construction such as *he ripped his collar open*.

Examples of particular sentential postitions are given below:

(15)

opened

(i) Postnominal

New lines opened and re-opened

The museum, opened 20 years ago

(ii) Prenominal

A *newly opened sex-shop*

The opened flowers

Three opened letters in her hand open

(iii) Postnominal

The choices open to everybody

(iv) Prenominal

Open account/admiration/air// open door/drawer/magazine

The Polish data are exemplified in the following constructions:

(16)

Podobne otwierane szafki lit. 'similar openable/being opened cupboards' (Mod) Prt (Mod) N

*Szlabany otwierane kart*ą 'bars opened with a card' NPrtMod

Otworzone gwałtownie drzwi 'a violently opened door' Prt Mod N

Drzwi otworzone zamaszyście 'door opened vigorously' NPrtMod

Otwarte linie kredytowe 'opened/open credit lines' Prt N Mod/ModN

Jej *otwarte*, jakby niewidzące oczy 'her open, as if blind, eyes' Prt Mod N

Notatki otwarte na niewłaściwej stronie 'notes opened on a wrong page' NPrt Mod

Tzw. *pytania otwarte* 'so-called open questions' NPrt [generic]

otwarty atak 'open attack'; *otwarte auta* 'open (convertible) cars'/*paleniska* 'open fire' PrtN [lexicalization] [attribute/state]

otwarte okna i drzwi Prt N [attribute/state] 'opened/open windows and doors'

When contrasted with the English (15), the Polish data (16) uncover the following patterns of the opening action phases:

(17)

Pol. *(i) otwierane – (ia) otwierając się - (ii) otworzone - (iii) otwarte*

Polish presents three past participial forms (i), (ii), (iii) and one present participial form (ia) as in (17) above:

(i) otwierane lit. 'being opened' conveys a sequential, processual onceptualization [not present in English in attributive position]; *Patient* reading *(ia)* present participial form (*otwierając się* – Medio-Passive, *otwierając – Agentive* reading);

(ii) otworzone li. 'having been opened'- sequential, terminative; and

(iii) otwarte-a 'opened, open'designates a state (change) with the final state profiled.

Ambiguity exists between (1) the participial sense {*drzwi zostały otwarte przez Tomka* 'the door was opened by Tom'} and (2) the adjectival sense {*drzwi są otwarte* 'the door is open'}.

Sense (1) presents a (participial, dynamic) profile with the terminative phase of the state change designated, whereas sense (2) profiles a stative, adjectival phase i.e., the final state alone.

The base Verbal form *otwierać* contains all the successive phases of the process [including the intitial negative state *–p* of the beginning of an action]; it can also denote an iterative (repetitive) action of opening. The base

Verbal form *otworzyć* profiles the perfective action of opening with the terminative phase of opening in focus.

The Event Structure frequencies in Polish are as follows:

(18)

EVENT [literal & metaphorical]

NKJP 300 mln

Y zamknięte 'closed' [state] //

X otwiera Y 'X is opening Y' [6,065]*– Y otwierane Z 'Y is being opened with Z' [202] [event procesual] → otwierane Y lit. 'being opened Y'[100]

Y [jest] otworzone' Y opened'[terminative phase, frequently postmodifier] [68] →

otworzone (modifier) Y [terminative attribute] 'opened Y'–> [33]

Y otwarte 'Y open' [4,700] / otwarte Y 'open Y' [4,200]] [final state]

As the consulted National Corpus of Polish (NKJP) is three times as big as the BNC, for the sake of a comparison the frequencies acquired for the Polish data should be normalized and divided into 3. The frequency of *otwart/l*,otworzył/l** as identified in the NKJP is *14,771*, which, when normalized to 100 million, gives ca 4,900 occurrences. Compared to the English *opened* with 28,502 occurrences in BNC, the frequency in English is over 5 times higher than in Polish. A language typological profile shows that the *modifiers* are more frequent and *metaphoric extensions* more numerous in English than in Polish.

7.3 Acceptability of prenominal past participle modifiers

While the Past (Passive) Participial modifier is acceptable in the prenominal position in Polish, in English the situation is more complex:

(19)

Oddaj zrobiony produkt lit. 'Return the made product' – 'Return the product you (or somebody else) made'.

The Polish NKJP corpus of 300 million units generates 400 Modifiers *zrobion** 'made', 50 in the postnominal position and 21 in the prenominal position. Here too the reason for the preference of the postnominal rather than the right-branching prenominal position may again involve processing rather than syntactic considerations, as Polish is a relatively free-word-order language.

A range of acceptable, semi-acceptable and acceptable structures in English, involving the attributive, past participle nominal modification position, cover the following cases:

(20)

?an eaten soup

?a drunk beer

?a seen accident

? a heard song

? a made/done product

versus

a half-made/partly made product (moth-eaten fur coat, *half-drunk beer*, etc.)

a well-done job

hot-spiced dish

and

a broken arm

a written statement

The prenominalization in English can be attributed to the status of (semantic) perfectivity of a given action, its completeness and boundedness. Therefore, *the change of state verbs* and verbs with additional *perfectivising modification* will be used as (attributive) prenominal participial modifiers. The reasons for this state of affairs are related to the perfectivity status of the verb and a corresponding participle. The perfective aspect is either *semantically inherent in the verbs*, as in the change of state verbs (*break*), or there is an imposed bounding, duration and permanence (*half-eaten sandwich*). Verbs such as *see* and *eat* are *semantically imperfective (unmarked) verbs [non-telic]* in English, which do not lend themselves to attributivisation. In the narrowly defined cases (*well-done*, *half-eaten*, *moth-eaten*), the perfectivity parameter, imposing mental bounding & possibly telos), is (more) clearly linguistically signalled.

The boundary acts similarly to the *state change*, which, in English, enables participles to be adjectivized in the pre-nominal position.

8. Language profiles and a research agenda for contrastive studies

On the basis of the exemplified qualitative and quantitative analyses, *individual language profiles* can be constructed, which act as *frames of reference*—tertia comparationis—in contrastive linguistics. What the contrastive

criteria involve is, therefore, both qualitative and quantitative in nature. Based on the *Event* structure, the qualitative criteria will acquire a more specific type from among a set of semantic and structural properties of different kinds. The quantitative criteria will foreground frequencies of items, patterns and clusters, as well as their collocational and distributional combinatorics. Taking as the point of departure the structural, procedural, and substantive tertia comparationis, together with the respective quantitative values, a *contrastive profile of languages* and language variety comparison can be constructed in cross-linguistic research.

The new research agenda for Contrastive Studies will thus invariably involve two levels of parameters, qualitative and quantitative, which will result in a *systematic procedure to contrast* the languages. It also makes it possible to carry out systematic *intralinguistic research* within one language[4]. The procedure considers the reconceptualization and approximation alignment between the systems, and eventually leads to a clearer identification of typological cross-linguistic and intra-systemic similarities and differences with far-reaching implications for translation studies and foreign-language education.

[4] e.g. contrasting reference corpus data with the internet Computer-Mediated Communication (CMC) as in the COST Action IS 0906 we are involved in on *Transforming Audiences, Transforming Societies*.

Acknowledgements

*I wish to express my thanks to Prof. I-wen Su (National Taiwan University) for providing the data and interpretation of the Chinese case.

**My thanks are due to Prof. Kingkarn Thepkanjana (Chulalongkorn University at Bangkok), for the examples, translation and advice on the materials in Thai.

References

Cristofaro, S. (2008) 'Asymmetric events, subordinations, and grammatical categories', in B. Lewandowska-Tomaszczyk (ed.), *Asymmetric events*, pp. 151-172, Amsterdam: J. Benjamins Pub.

Davidson, D. (1980) *Essays on action and events*, Oxford: Clarendon Press.

Dziwirek, K. & Lewandowska-Tomaszczyk, B. (2010) *Complex emotions and grammatical mismatches,* Berlin: Mouton de Gryuter.

Gardenförs, P. (2004) *Conceptual spaces: The geometry of thought,* Cambridge: The MIT Press.

Hauser, M. D., Chomsky, N. & Fitch, T. (2002) The faculty of language: What is it, who has it, and how did it evolve, *Science*, vol. 298, pp. 1569-1579

Lakoff, G. (1987) *Women, fire and dangerous things: What categories reveal about human mind*, Chicago: University of Chicago Press.

Langacker, R. W. (1987-1991) *Foundations of cognitive grammar,* vols. 1-2, Stanford: Stanford University Press.

Lewandowska-Tomaszczyk, B. (1987) *Conceptual analysis, linguistic meaning and verbal interaction,* Łódź, Poland: Łódź University Press.

Lewandowska-Tomaszczyk, B. (1996) *The depth of negation-cognitive-semantic study*, Łódź, Poland: Łódź University Press.

Lewandowska-Tomaszczyk, B. (1999) 'A cognitive-interactional models of cross-linguistic analysis: New perspectives on 'tertium comparationis' and the concept of equivalence', in B. Lewandowska-Tomaszczyk (ed.), *Cognitive perspectives on language*, Frankfurt: Peter Lang.

Lewandowska-Tomaszczyk, B. (2007) 'Polysemy, prototypes, and radial categories', in D. Geeraerts & H. Cuyckens (eds.)*, The Oxford handbook of cognitive linguistics,* Oxford: Oxford University Press.

Lewandowska-Tomaszczyk, B. (2008) 'Asymmetries in participial modification'*,* in B. Lewandowska-Tomaszczyk (ed.), *Asymmetric Events*, Amsterdam: Benjamins.

Lewandowska-Tomaszczyk, B. (2010) 'Re-conceptualization and the emergence of discourse meaning as a theory of translation', in B. Lewandowska-Tomaszczyk & M. Thelen (eds.), *Meaning in translation,* Frankfurt: Peter Lang.

Lewandowska-Tomaszczyk, B. (2011) 'Events as they are', in P. Stalmaszczyk (ed.), *Turning points in the philosophy of language and linguistics.* Frankfurt: Peter Lang.

Lewandowska-Tomaszczyk, B. (2012) 'Explicit and tacit: An interplay of the quantitative and qualitative approaches to translation', in M. P. Oakes & J. Meng (eds.), *Quantitative methods in corpus-based translation studies: A practical guide to descriptive translation research*, Amsterdam: John Benjamins Publishing.

Lewandowska-Tomaszczyk, B. (2012) *Approximativeness and tolerance threshold in communication*, Poland: Lodz University.

Lewandowska-Tomaszczyk, B. (ed.). (2008) *Asymmetric events,* Amsterdam: John Benjamins.

Lewandowska-Tomaszczyk, B. & Dziwirek, K. (eds.). (2009) *Studies in cognitive corpus linguistics,* Frankfurt: Peter Lang.

Lewandowska-Tomaszczyk, B. & Wilson, P. (2011) 'Culture-based conceptions of emotion in Polish and English', in S. Goźdź-Roszkowski (ed.), *Explorations across languages and corpora*, Frankfurt: Peter Lang.

Progovac, L. (2010) Syntax: Its evolution and its representation in the brain, *Biolinguistics,* vol. 4, pp. 234-254.

Seuren, P. (1985) *Discourse semantics,* Oxford: Blackwell.

Talmy, L. L. (1985) 'Patterns: Semantic structure in lexical forms', in T. Shopen (ed.), *Language typology and syntactic description*, vol. 3, Cambridge: Cambridge University Press.

Vendler, Z. (1957) Verbs and times, *Philosophical Review*, vol. 66, pp. 143-60.

Verkuyl, H. A. (1993) *A theory of aspectuality: The interaction between temporal and atemporal structure,* Cambridge: Cambridge University Press.

Wright, V. & Georg, H. (1963) *Norm and action: A logical inquiry*, London: Routledge and Kegan Paul.

Zacks, J. M. & Tversky, B. (2001) Event structure in perception and conception, *Psychological Bulletin*, vol. 127, January, pp. 3-21.

Corpora

British National Corpus (BNC)

National Corpus of Polish (nkjp.pl) *Narodowy Korpus jezyka Angielskiego (NKJP)*

Samplers

Longman & Microconcord (15-million segments of English)

PELCRA (10- and 20-million segments of Polish)