

# Testing the Annotation Consistency of Hallidayan Transitivity Processes: A Multi-variable Structural Approach

Min Dong<sup>1</sup>, Xiaoyan Liu<sup>1</sup>, Alex Chengyu Fang<sup>2,3</sup>

<sup>1</sup>School of Foreign Languages, Beihang University, PR China

<sup>2</sup>Department of Linguistics and Translation, City University of Hong Kong, HKSAR

<sup>3</sup>The Halliday Centre for Intelligent Applications of Language Studies, City University of Hong Kong, HKSAR  
{mdong, liuxiaoyan0112}@buaa.edu.cn, acfang@cityu.edu.hk

## Abstract

SFL seeks to explain identifiable, observable phenomena of language use in context through the application of a theoretical framework which models language as a functional, meaning making system (Halliday & Matthiessen 2004). Due to the lack of explicit annotation criteria and the divide between conceptual vs. syntactic criteria in practice, it has been a tough job to achieve consistency in the annotation of Hallidayan transitivity processes. The present study proposed that explicit structural and syntactic criteria should be adopted as a basis. Drawing on syntactic and grammatical features as judgement cues, we applied structurally oriented criteria for the annotation of the process categories and participant roles combining a set of interrelated syntactic variables and established the annotation criteria for contextualised circumstantial categories in structural as well as semantic terms. An experiment was carried out to test the usefulness of these annotation criteria, applying percent agreement and Cohen's kappa as measurements of interrater reliability between the two annotators in each of the five pairs. The results verified our assumptions, albeit rather mildly, and, more significantly, offered some first empirical indications about the practical consistency of transitivity analysis in SFL. In the future work, the research team expect to draw on the insights and experience from some of the ISO standards devoted to semantic annotation such as dialogue acts (Bunt et al. 2012) and semantic roles (ISO-24617-4, 2014).

**Keywords:** annotation consistency, Hallidayan transitivity process, structurally oriented criteria

## 1. Introduction

Systemic functional linguistics (SFL) founded by Halliday (1967/8), together with American “West-Coast” functionalism represented by Hopper & Thompson (1980), has “moved up” its descriptions of transitivity from the rank of the verb to that of the clause and thus enabled more holistic and explanatory accounts of transitivity. From the vantage point of semantics, significantly, SFL characterises transitivity in functional terms as a lexicogrammatical resource for construing a quantum of change in our experience of the flow of events around us and inside us. As a result, the domain of transitivity is modelled as a configuration of a process (a verb group), participants directly involved in it (nominal groups) and attendant circumstances (adverbs and prepositional phrases). Six process categories are proposed in the grammar of SFL, among which, in terms of relative frequency, material, relational, mental and verbal processes constitute the major types while behavioural and existential processes make up the two minor categories (Matthiessen 1999, 2014). There is a general agreement that transitivity analysis is problematic in practice (O'Donnell et al 2009). The community survey about the annotation practice of transitivity in O'Donnell et al (2009) reveals that the divide between the use of syntactic (based on the syntactic structure of the clause) vs. conceptual (on the basis of the underlying action or event represented by the clause) criteria is widespread throughout the SFL community. The analysts, however, may unconsciously find themselves forced to fall back on purely semantic criteria, particularly when encountered with

challenging cases in which lexicogrammatical structures associated with one process type appear to be realising a different one in terms of the meaning being expressed (O'Donnell et al 2009). Explicit statements of annotation criteria of transitivity are needed for determining how each and every clause should be annotated in terms of process configuration, including one of the six processes, the obligatory or optional participants and attendant circumstances.

Despite the fact that the descriptions offered in Halliday's accounts of transitivity are primarily based on semantic criteria, certain key grammatical criteria for transitivity categorisation (such as preferred tense/aspect, and the potential to project) have been elaborated (Halliday 1994: 115-16). Essentially, all possible clausal configurations including subject types, verb forms and complementation patterns need to be taken into account. It seems more practical to view the issue in terms of syntactic variables, which may (or may not) come into play in deciding particular transitivity categories, and which may combine a wide range of judgements. We are thus motivated to propose an approach which draws on a wider array of syntactic variables for automated grammatical analysis from Fang (2007) to establish explicitly stated sets of criteria for the annotation of transitivity. It is expected to be a promising approach to pinning down more precisely than has been the case so far with the kinds of variables that are relevant for particular verbs or groups of verbs (or of verb senses). To test these assumptions, the present study carries out an experiment by giving ten postgraduates a set of explicit annotation guides (to

be expounded in Section 2) for their analysis of the component sentences of a news article randomly assigned to each of them.

## 2. Structurally Oriented Annotation of Hallidayan Transitivity Categories

According to Matthiessen (1999), a principle can be held that the more frequently a process type is selected in text, the more highly elaborated it is systemically in terms of a larger number of verb classes, verbs and thus verb senses that can be assigned to it. Based on Levin's (1993) verb classes, Matthiessen (2014) further reveals that the material process is most highly elaborated in lexis and is followed by the two highly elaborated types of

mental and verbal processes; the two least frequent types of behavioural and existential processes are accordingly least elaborated. An exception is found in "relational" process clauses: while they are equally frequent compared with "material" ones, they are far less highly elaborated, attributable to the characteristic that the lexical elaboration is allocated to the participants of the clause rather than to the process verbs. According to such a descending order of the six process types in terms of relative elaboratedness in lexis, a descriptive framework of transitivity is devised below as a basis of the coding criteria in syntactic terms for the annotation of transitivity categories. First of all, Table 1 offers descriptions of each of the transitivity terms in each of the six process types.

<i>Process type</i>	<i>Process category &amp; Participant role</i>	<i>Description</i>
Material	Material process	creative (event/ thing) & transformative experience
	Actor	the role responsible for bringing about a change
	Goal	the entity that is brought into existence
	Beneficiary	the role who is given goods or for whom a service is performed
	Scope	the role which construes the process itself or the domain over which the process takes place
	Initiator	the role responsible for making the Actor perform an action
Mental	Mental process	cognitive, desiderative, emotive & perceptive experience
	Senser	the conscious role who thinks, perceives, feels or desires something or someone
	Phenomenon	the role that reflects what is thought, perceived, felt or desired
	Inducer	the role that causes the Senser to think, perceive, feel or desire something
Verbal	Verbal process	non-projecting (communicating & targeting) & projecting (imperating & indicating) experience
	Sayer	the role who puts out a signal of symbolic exchange of meaning
	Verbiage	the role which denotes the content or nature of the message itself
	Receiver	the role to whom the message is addressed
	Target	the role usually in the verbal clause of judgement, such as praise, blame, criticism, representing the entity that is the object of judgement by the Sayer
Relational: Attributive	Relational process	circumstantial (causal/comparative/locative/matter), intensive (attributed/ non-attributed) & possessive (benefactive/ non-benefactive) relations
	Carrier	the entity to which the Attribute is ascribed
	Attribute	the class to which the Carrier is attributed
	Attributor	the role that brings about the attribution of the Attribute to the Carrier
Relational: Identifying	Relational process	circumstantial (causal/comparative/locative/matter), intensive (assigned/ non-assigned) & possessive (benefactive/ non-benefactive) relations
	Token	the specific embodiment which is assigned to a more generalisable category of Value in the relational clause of identification
	Value	the more general category which is assigned to a specific realisation of Token in the relational clause of identification
	Assigner	the role which assigns the relationship of identity between the Token and the Value
Behavioural	Behavioural process	inter-active & intro-active (conscious activity/ physiological) experience
	Behaver	a conscious being inherent in the process of physiological or psychological behaviour
Existential	Existential process	entity & event existence
	Existent	an entity existing in concrete or abstract space, or an event occurring in time

Table 1: Description of process categories and participant roles in the six process types

Halliday (1994) stresses that in order to posit a grammatical category there must be a "lexico-grammatical reflex of the difference [in meaning]". This all clearly suggests that the focus is the lexicon and that the structural configurations are always associated with particular meanings. In line with a corpus-based approach to syntactic analysis, Fang (2007) demonstrates how

automated grammar analysis is able to be implemented using AUTASYS (Fang 1996) and Survey Parser (Fang 2006). On this basis, we draw on a range of 20 grammatical and syntactic variables from Fang (2007) as shown in Table 2. The structurally oriented coding criteria for transitivity annotation are derived accordingly.

Process	Transitivity	Verb Complementation Types						Syntactic Functions and Structural Features												
		cop	cxtr	ditr	intr	montr	trans	A: phr	by	CO	CS	CT	EX	OD: cl	OD: ing	OD: phr	OD: to	OI	SU	SU: agent
Material	Material pro		√	√	√	√	√						√		√		√			
	Actor								√										√	
	Goal								√										√	
	Beneficiary																	√	√	
	Scope							√		√						√				
	Initiator																			
Mental	Mental pro		√			√	√						√	√			√			
	Senser								√										√	
	Phenomenon								√										√	
	Inducer																			√
Verbal	Verbal pro		√	√	√	√	√						√	√			√			
	Sayer								√										√	
	Verbiage								√										√	
	Receiver																	√	√	
	Target								√										√	
Relational: Attributive	Relational pro	√				√								√			√			
	Carrier																		√	
	Attribute								√		√									
	Attributor																			√
Relational: Identifying	Relational pro	√				√								√			√			
	Token																		√	
	Value								√		√								√	
	Assigner																			√
Behavioural	Behavioural pro				√									√			√			
	Behaver																		√	
Existential	Existential pro				√							√		√			√			
	Existent																		√	

Table 2: Grammatical and syntactic variables (Fang 2007) associated with configured transitivity categories in each of the six process types

cop – copula verb  
cxtr – complex transitive verb  
ditr – ditransitive verb  
intr – intransitive verb  
montr – monotransitive verb  
trans – transitive verb  
A:phr – verb-preposition-adverbial phrase  
by – Prepositional complement of *by*-phrase,  
CO – Object complement

CS – Subject complement  
CT – Transitive complement  
EX – Existential *there* structure  
OD – Direct object  
OD:cl – Finite *that*-clause as object  
OD:ing – Non-finite *ing*-clause as object  
OD:phr – Verb-noun-object phrase  
OD:to – Non-finite *to*-infinitive clause as object  
OI – Indirect object

SU – Subject  
SU:agent – Agentive subject

Concerning circumstance categories, in terms of the frequency and size of prepositions used, as observed in Matthiessen (1999), “location” is both the most frequently used option and also the most highly elaborated, followed by “manner”, “cause” and “accompaniment”. Interestingly, while “extent” is more frequently used than “matter”, it is much less elaborated than the latter, which is possibly related to the fact that “extent” is typically frequently realised by adverbial groups. Table 3 presents a descriptive framework of these circumstantial terms in descending order of lexical elaboration, both conceptually and syntactically. It is important to note that circumstantial elements perform the function of adverbials. The coding criteria for the annotation of circumstantial categories are therefore established in structural as well as semantic terms.

Type	Subtype	Grammatical category	Syntactic functions
Location	Place	PP/ AVP	Adverbial
	Time	PP/ AVP	Adverbial
Manner	Means	PP/ AVP	Adverbial
	Quality	PP/ AVP	Adverbial
	Comparison	PP/ AVP	Adverbial
	Degree	PP/ AVP	Adverbial
Cause	Reason	PP	Adverbial
	Purpose	PP	Adverbial
	Behalf	PP	Adverbial
Accompaniment	Comitative	PP/ AVP	Adverbial
	Additive	PP/ AVP	Adverbial
Extent	Distance	PP/ AVP	Adverbial
	Duration	PP/ AVP	Adverbial
	Frequency	AVP	Adverbial
Matter		PP	Adverbial
Contingency	Condition	PP	Adverbial
	Default	PP	Adverbial
	Concession	PP	Adverbial
Role	G guise	PP	Adverbial
	Product	PP	Adverbial
Angle	Source	PP	Adverbial
	Viewpoint	PP	Adverbial

Table 3: Circumstantial categories in semantic, grammatical and syntactic terms (Halliday & Matthiessen 2004)

### 3. Experiment on Transitivity Annotation

In this section we describe an experiment carried out on transitivity annotation. It is based on a text of news report which comprises forty sentences in total. Ten postgraduates in the field of linguistics were recruited and assigned into five pairs of two annotators. Each pair were given eight sentences randomly selected from the news report. They were all instructed and trained about the structurally oriented annotation criteria in terms of interrelated grammatical and syntactic variables for the annotation of process categories and participant roles in each of the six process types and attendant circumstantial categories. More specifically, each

annotator was instructed to analyse the test sentences according to the following requirements:

1. All directly embedded clausal constituents which perform sentential syntactic functions, whether finite or non-finite, are required to undertake transitivity annotation.
2. All clausal constituents indirectly embedded in nominal, adjectival or prepositional phrases, whether finite or non-finite, should be exempted from transitivity annotation.

In order to determine interannotator agreement between the two coders in each of the five pairs and the general level of agreement among the ten coders with regard to their transitivity annotation of the whole text, two methods of measurement of interrater reliability are deployed in the present study, namely, Percent agreement and Cohen’s kappa (McHugh 2012):

Percent agreement. This statistic is calculated by dividing the number of zero difference codings by the number of variables provides a measure of percent agreement between the raters. It is also directly interpretable as the percent of data that are correct. It is typically recommended that 80% agreement as the minimum acceptable interrater agreement.

Cohen’s kappa. In view of the limitation of Percent agreement that it does not consider the possibility that raters guessed on scores and may thus overestimate the true agreement among raters, the Kappa was designed to take account of the possibility of random guesses (Cohen 1960). The calculation of Cohen’s kappa may be performed according to the following formula:

$$k = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)}$$

where  $\Pr(a)$  represents the actual observed agreement, and  $\Pr(e)$  represents the expected chance agreement. Notably the sample size consists of the number of observations made across which raters are compared. Kappa is a form of correlation coefficient based on the *chi*-square table. While correlation coefficients cannot be directly interpreted, a squared correlation coefficient is directly interpretable, namely the amount of variation in the dependent variable that can be explained by the independent variable. The estimate of such variance accounted for is usually obtained by squaring the correlation value. It is by extension of this logic that the calculation of  $\Pr(e)$ , the chance agreement is carried out by squaring the amount of accuracy in the data due to congruence among the raters.

It is noted that Kappa is typically a considerable reduction in the level of congruence compared with percent agreement, and thus it has the limitation that it may lower the estimate of agreement greatly. The greater the expected chance agreement, the lower the resulting value of the Kappa. In addition, it

cannot be directly interpreted. However, any Kappa value below 0.60 indicates inadequate agreement among the raters and little confidence should be placed in the study results. Considering the great chance of guessing in semantic annotation like transitivity annotation focused on in this study, and the fact that the ten annotators have been trained on the annotation criteria proposed in structural terms, we are thus motivated to calculate both percent agreement and Kappa.

#### 4. Results and Discussion

Using the two statistical techniques of Percent agreement and Cohen’s kappa, Table 4 shows the calculation results of measurement of interrater reliability between the two transitivity annotators in each of the five pairs. As displayed, the scores of Percent agreement indicate that four pairs achieve a fairly adequate agreement between the two annotators in each of them. Pair 4, however, performs as an outlier, far below 80% agreement. A closer look at the Kappa values, however, shows that the interrater reliability accomplished by the two annotators in the five pairs visibly diverge among themselves, constituting a typical cline ranging from “None” and “Weak” at one end to “Moderate” and “Strong” at the other. More interestingly, Pair 3 and Pair 5 both gain notably adequate agreement, namely well above 80% agreement, and accordingly they respectively achieve moderate and strong agreement in terms of Kappa value. This observation might bring additional support for the compatibility and complementarity between the two techniques as discussed in the literature (McHugh 2012).

	Percent	Kappa	Level of Kappa agreement	% of reliable data
Pair 1	0.74	0.41	Weak	about 16%
Pair 2	0.76	0.16	None	about 3%
Pair 3	0.86	0.61	Moderate	about 36%
Pair 4	0.63	0.12	None	about 2%
Pair 5	0.94	0.80	Strong	about 64%

Table 4: Interrater reliability in Percent agreement and Cohen’s kappa

With respect to the transitivity annotation of the whole text of the selected news article, it can be suggested that each of the ten annotators made a deliberate choice of an annotation of process categories, participant roles and circumstance types, and that the majority have had a certain amount of correct understanding of the structurally oriented annotation criteria proposed in this study. The detailed information of the distribution of correct (“normal”), incorrect (“abnormal”) and incongruent (“normal+ abnormal”) scores between the two annotators in each of the five pairs is offered by Table 5. It is observed that the percentage of data that both Annotator A and Annotator B make correct (“normal”) judgments on is all above 50%

across the five pairs, with Pair 4 obtaining the lowest and Pair 5 the highest.

<i>Pair 1</i>					
		Annotator B			
		Normal	Abnormal		
Annotator A	Normal	63	15	78	rm1
	Abnormal	15	23	38	rm2
		78	38	116	
		cm1	cm2	n	
<i>Pair 2</i>					
		Annotator B			
		Normal	Abnormal		
Annotator A	Normal	57	13	70	rm1
	Abnormal	6	4	10	rm2
		63	17	80	
		cm1	cm2	n	
<i>Pair 3</i>					
		Annotator B			
		Normal	Abnormal		
Annotator A	Normal	48	4	52	rm1
	Abnormal	6	12	18	rm2
		54	16	70	
		cm1	cm2	n	
<i>Pair 4</i>					
		Annotator B			
		Normal	Abnormal		
Annotator A	Normal	46	22	68	rm1
	Abnormal	10	9	19	rm2
		56	31	87	
		cm1	cm2	n	
<i>Pair 5</i>					
		Annotator B			
		Normal	Abnormal		
Annotator A	Normal	67	1	68	rm1
	Abnormal	4	13	17	rm2
		71	14	85	
		cm1	cm2	n	

Table 5: Correct, incorrect and incongruent annotations

Now we draw attention to the distribution of the annotations of the six process types and the nine circumstance categories across the three types of interannotator agreement, including both correct (“both normal”), both incorrect (“both abnormal”) and correct+incorrect (“normal+ abnormal”), in each of the five pairs. In this way we are enabled to determine to what extent students have learnt annotating these transitivity categories reliably guided by the structurally oriented annotation criteria expounded in Section 2. As shown below, while Table 6 provides the basic information of distribution of the three types of interannotator agreement of the six process types across each of the five pairs, Table 7 presents the parallel distributional information of the nine circumstantial categories.

<i>Pair 1</i>	Material Pro	Relational Pro: Attributive	Relational Pro: Identifying	Mental Pro	Verbal Pro	Existential Pro	<i>Sub-total</i>
<i>Both Normal</i>	31.58%	2.63%	2.63%	5.26%	2.63%	0.00%	44.74%
<i>Both Abnormal</i>	15.79%	2.63%	2.63%	2.63%	0.00%	0.00%	23.68%
<i>Normal+Abnormal</i>	18.42%	2.63%	5.26%	0.00%	2.63%	2.63%	31.58%
<i>Total</i>	65.79%	7.89%	10.53%	7.89%	5.26%	2.63%	100.00%
<i>Pair 2</i>	Material Pro	Relational Pro: Attributive	Relational Pro: Identifying	Mental Pro	Verbal Pro	Existential Pro	<i>Sub-total</i>
<i>Both Normal</i>	22.22%	18.52%	7.41%	0.00%	14.81%	7.41%	70.37%
<i>Both Abnormal</i>	0.00%	3.70%	0.00%	0.00%	0.00%	0.00%	3.70%
<i>Normal+Abnormal</i>	11.11%	3.70%	3.70%	3.70%	3.70%	0.00%	25.93%
<i>Total</i>	33.33%	25.93%	11.11%	3.70%	18.52%	7.41%	100.00%
<i>Pair 3</i>	Material Pro	Relational Pro: Attributive	Relational Pro: Identifying	Verbal Pro	<i>Sub-total</i>		
<i>Both Normal</i>	18.18%	13.64%	0.00%	18.18%	50.00%		
<i>Both Abnormal</i>	18.18%	4.55%	4.55%	4.55%	31.82%		
<i>Normal+Abnormal</i>	4.55%	4.55%	0.00%	9.09%	18.18%		
<i>Total</i>	40.91%	22.73%	4.55%	31.82%	100.00%		
<i>Pair 4</i>	Material Pro	Relational Pro: Attributive	Mental Pro	Verbal Pro	<i>Sub-total</i>		
<i>Both Normal</i>	26.32%	0.00%	0.00%	21.05%	47.37%		
<i>Both Abnormal</i>	5.26%	5.26%	5.26%	0.00%	15.79%		
<i>Normal+Abnormal</i>	21.05%	0.00%	0.00%	15.79%	36.84%		
<i>Total</i>	52.63%	5.26%	5.26%	36.84%	100.00%		
<i>Pair 5</i>	Material Pro	Relational Pro: Attributive	Mental Pro	Verbal Pro	Behavioral Pro	<i>Sub-total</i>	
<i>Both Normal</i>	37.93%	3.45%	10.34%	27.59%	0.00%	79.31%	
<i>Both Abnormal</i>	3.45%	6.90%	3.45%	0.00%	3.45%	17.24%	
<i>Normal+Abnormal</i>	0.00%	3.45%	0.00%	0.00%	0.00%	3.45%	
<i>Total</i>	41.38%	13.79%	13.79%	27.59%	3.45%	100.00%	

Table 6: Correct, incorrect and incongruent annotations of different processes

Pairwise, it is observable that the two coders in Pair 5 exhibit a high level of correctness and agreement in judgments of the process types concerned, whereas Pair 3, Pair 1 and Pair 4 perform considerably weaker due to a saliently larger proportion of data judged either as both incorrect or as incongruent with one of them being incorrect. More notably, close to Pair 5, the two annotators in Pair 2 also achieve a high level of correctness in their annotations, but they are far more remarkably at odds with each other, which explains their low Kappa value. This finding verifies the usefulness of the instruction of the structurally-oriented coding criteria proposed in this study. It also demonstrates the necessity of addressing the issue of interrater reliability encountered in transitivity annotation due to the lack of more fully explicit annotation criteria in structural rather than semantic terms and the long-standing divide between the use of conceptual as opposed to syntactic criteria in the community. Process-wise, it seems that the problems with annotation of the six process types are evenly distributed across each of the five pairs, regardless of their larger or smaller relative frequency and elaboratedness in lexis. Concerning the annotation of circumstantial categories, however, Table 7 presents a visibly different picture.

Notably, a high level of correctness and agreement in judgments of the five circumstantial categories in Pair 5 is not as saliently exhibited as in Table 6. In addition, the two annotators in Pair 3, perform slightly better than Pair 5 as they achieve a narrowly higher level of correctness and a broadly lower level of incorrectness in their annotations, though exhibiting a larger data of incongruence. This observation may be related to the far smaller number of different types of circumstance to be annotated in Pair 3. Furthermore, compared with the parallel data of annotations of different process types offered in Table 6, while Pair 1 and Pair 4 make a similarly substantially poor performance both in terms of a higher level of incorrectness and of incongruence, Pair 2 unexpectedly achieves a much lower level of correctness and agreement in their annotations of circumstantial categories. This result might suggest the urgency of developing explicit syntactic criteria for transitivity annotation given that the instances where prepositional phrases function as adverbials are often easily confused with those in which they function as qualifiers of nominal or adjectival phrases. Circumstance-wise, quite evidently, the annotation of each of the nine circumstantial categories is problematic at a general level, whether they are of larger or smaller relative frequency and lexical elaboration.

<i>Pair 1</i>	Cir: Location	Cir: Manner	Cir: Cause	Cir: Accompaniment	Cir: Matter	<i>Sub-Total</i>			
<i>Both Normal</i>	31.58%	15.79%	0.00%	0.00%	0.00%	47.37%			
<i>Both Abnormal</i>	0.00%	10.53%	0.00%	0.00%	0.00%	10.53%			
<i>Normal+Abnormal</i>	21.05%	5.26%	5.26%	5.26%	5.26%	42.11%			
<i>Total</i>	52.63%	31.58%	5.26%	5.26%	5.26%	100.00%			
<i>Pair 2</i>	Cir: Location	Cir: Manner	Cir: Role	Cir: Matter	Cir: Angle	<i>Sub-Total</i>			
<i>Both Normal</i>	28.57%	7.14%	0.00%	0.00%	0.00%	35.71%			
<i>Both Abnormal</i>	7.14%	14.29%	0.00%	0.00%	0.00%	21.43%			
<i>Normal+Abnormal</i>	7.14%	14.29%	7.14%	7.14%	7.14%	42.86%			
<i>Total</i>	42.86%	35.71%	7.14%	7.14%	7.14%	100.00%			
<i>Pair 3</i>	Cir: Location	Cir: Manner	Cir: Angle	<i>Sub-Total</i>					
<i>Both Normal</i>	38.46%	7.69%	15.38%	61.54%					
<i>Both Abnormal</i>	7.69%	7.69%	0.00%	15.38%					
<i>Normal+Abnormal</i>	23.08%	0.00%	0.00%	23.08%					
<i>Total</i>	69.23%	15.38%	15.38%	100.00%					
<i>Pair 4</i>	Cir: Location	Cir: Extent	Cir:Manner	Cir: Cause	Cir: Contingency	Cir: Accompaniment	Cir: Matter	Cir: Angle	<i>Sub-Total</i>
<i>Both Normal</i>	26.92%	0.00%	3.85%	3.85%	0.00%	0.00%	0.00%	0.00%	34.62%
<i>Both Abnormal</i>	7.69%	0.00%	3.85%	3.85%	0.00%	0.00%	0.00%	3.85%	19.23%
<i>Normal+Abnormal</i>	7.69%	7.69%	7.69%	0.00%	3.85%	3.85%	11.54%	3.85%	46.15%
<i>Total</i>	42.31%	7.69%	15.38%	7.69%	3.85%	3.85%	11.54%	7.69%	100.00%
<i>Pair 5</i>	Cir: Location	Cir: Extent	Cir:Manner	Cir: Cause	Cir: Contingency	<i>Sub-Total</i>			
<i>Both Normal</i>	42.86%	7.14%	0.00%	7.14%	0.00%	57.14%			
<i>Both Abnormal</i>	0.00%	7.14%	14.29%	0.00%	7.14%	28.57%			
<i>Normal+Abnormal</i>	7.14%	0.00%	7.14%	0.00%	0.00%	14.29%			
<i>Total</i>	50.00%	14.29%	21.43%	7.14%	7.14%	100.00%			

Table 7 Correct, incorrect and incongruent annotations of different circumstantial categories

## 5. Conclusion

Tough challenges have been encountered for the task of transitivity annotation due to the stratal blurring of whether transitivity is positioned at the level of semantics or lexico-grammar in SFL. The situation is aggravated by the lack of explicit annotation criteria as well as the divide between conceptual vs. syntactic criteria in practice. To help address these issues, the present study proposed that explicit structural and syntactic criteria should be adopted as a basis to maximize the consistency in annotations that are conceptual and semantic in nature. Drawing on syntactic and grammatical features as judgement cues, we applied structurally oriented criteria for the annotation of the Hallidayan process categories and participant roles combining a set of interrelated syntactic variables and established the annotation criteria for contextualised circumstantial categories in structural as well as semantic terms. The experiment, which was carried out to test the usefulness of these annotation criteria, applied percent agreement and Cohen's kappa as measurements of interrater reliability between the annotators in each of the five pairs. The results verified our assumptions, albeit rather mildly, and, more significantly, offered some first empirical indications about the practical consistency of

transitivity analysis in SFL. The results have also produced insights and suggestions for some future work. The research team expect to integrate automated syntactic analysis and manual transitivity annotation to compute the probability of the association between the grammatical and syntactic categories and transitivity categories in preparation for the implementation of automated transitivity analysis. It is also expected to draw on the insights and experience from some of the ISO standards devoted to semantic annotation such as dialogue acts (Bunt et al. 2012) and semantic roles (ISO-24617-4, 2014).

## 6. Acknowledgements

Research described in this article was partially supported by grants received from Beijing Municipality Planning Office of Philosophy and Social Sciences (Project No 19YYA001) and CityU Strategic Interdisciplinary Research Grant (Project No 7020036).

## 7. Bibliographical References

Bunt, H, Alexandersson, J, Choe, J W, Fang, A C, Hasida, K, Petukhova, V, Popescu-Belis, A, & Traum, D (2012). *ISO 24617-2:2012 Language resource management – Semantic annotation*

- framework (SemAF) – Part 2: Dialogue acts.*  
Geneva: The International Organization for Standardization.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20: 37-46.
- Fang, A.C. (1996). AUTASYS: Automatic Tagging and Cross-Tagset Mapping. In S. Greenbaum (Ed.), *Comparing English World Wide: The International Corpus of English*. Oxford: Oxford University Press, pp. 110--124.
- Fang, A.C. (2006). Evaluating the Performance of the Survey Parser with the NIST Scheme. In A. Gelbukh (Ed.), *LNCS 3878: Computational Linguistics and Intelligent Text Processing*. Berlin Heidelberg: Springer-Verlag, pp. 168--179.
- Fang, A.C. (2007). *English Corpora and Automated Grammatical Analysis*. Beijing: The Commercial Press.
- Halliday, M, A, K. (1994). *An Introduction to Functional Grammar*. London: Edward Arnold, 2nd edition.
- Halliday, M.A.K. (1967/8). Notes on transitivity and theme in English. *Journal of Linguistics*, 3(1): 37-81, 3(2): 199-244, 4(2): 179-215.
- Hopper, P. & Thompson, S.A. (1980). Transitivity in grammar and discourse. *Language*, 56(2): 251-99.
- ISO-24617-4. 2014. *Language resource management - Semantic annotation framework (SemAF) - Part 4: Semantic roles (SemAF-SR)*. Geneva: International Organization for Standardization.
- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.
- Matthiessen, C. M. I. M. (1999). The system of transitivity: An exploratory study of text-based profiles. *Functions of Language*, 6(1): 1-51.
- Matthiessen, C. M. I. M. (2014). Extending the description of process type within the system of transitivity in delicacy based on Levinian verb classes. *Functions of Language*, 21(2): 139-175.
- McHugh, M.L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3): 276-282.
- O'Donnell, M., Zappvigna, M. and Whitelaw, C. (2009) A survey of process type classification over difficult cases. In J. Carys and E. Ventola (Eds.), *From Language to Multimodality: New Developments in the Study of Ideational Meaning*. Bristol: Equinox Publishing, pp. 47--64.