From Verb to Action: A Mechanics-based Approach for Trajectory-Pattern Interpretation of Manipulation Verbs

Man Shi¹ and Hongbo Wang²

¹School of International Studies, Shaanxi Normal University, 199 South Chang’an Road, Xi’an, China
²School of Mechanical Engineering, Northwestern Polytechnical University, 127 West Youyi Road, Xi’an, China
Email: wanghongbo@nwpu.edu.cn

Abstract. Interpreting manipulation verbs in natural-language instructions into manipulation actions is an indispensable ability for future service robots. As a typical example of manipulation verbs, cutting and breaking (C&B) verbs has gained increasing attention, and various interpretation methods have been proposed in both fields of robotics and natural language processing. However, the existing methods are either relying on the manual meaning matching with tools, objects, and demonstration videos, or based on qualitative and empirical knowledge. Consequently, it is unable to quantitatively interpret the C&B verbs into the robots’ trajectory patterns, which are the necessary information for performing manipulation actions. To address this problem, in this paper, we propose a mechanics-based approach, using the contact, twist, and wrench formulations, for interpreting the C&B verbs into quantitative, computable, and verifiable forms. The proposed method leads to the quantitative constraints and unified validation criteria for robot trajectory generation of manipulation tasks, as well as the semantic analysis in natural language processing.

1. Introduction

Turning natural-language instructions into actions is one of the key features of future service robots and human–computer interfaces [1]. Many investigations have been directed to interpret various types of verb instructions into their corresponding actions: e.g., route following [2] and target location inference [3]. More recently, the interpretation of manipulation verbs is gaining increasing attention [4][5], as the expected translation framework has the potential to improve robots’ manipulation abilities.

Cutting and breaking (C&B) verbs, which is a typical class of manipulation verbs and corresponds to the separations in the material integrity of objects [6], are the fundamental and day-to-day actions that have been central to hominid cognition and culture for more than two million years [7]. Accordingly, C&B verbs appear in many languages. For example, the monosyllabic Mandarin Chinese verbs 切 (qiē), 剁 (duò), 砍 (kǎn), 割 (gē), 剁 (xiāo), 割 (guō), 勒 (pǐ), 锯 (jù), 撕 (sī), 拆 (chē), 折 (zhé), and 押 (bāi); the English verbs cut, slice, shear, hack, chop, mow, pare, whistle, peel, shave, scrape, scratch, cleave, and Japanese verbs 切る (kiru), 刻む (kizamu), 裂る (kiri), 割る (karu), 剃る (soru), 割る (waru), and 削る (kezuru), etc.

Consequently, in natural language instructions, the C&B verbs can be frequently encountered in both industrial and domestic scenarios. For a robot that is capable of performing C&B actions, it is necessary to interpret the verbs into the trajectory patterns of the robots’ end-effectors. However, the verb meanings are currently either predefined for a few designated tools and objects [4], or obtained from...
objects’ state changes via demonstration videos [4], whereas the verbs’ physical meanings, such as trajectory patterns, are left unstudied.

Besides robotics and human-computer interaction, C&B verbs have also been intensively investigated in linguistics for its important role in categorization [8-10] and syntax-semantics interface [11][12], while the meanings of the C&B verbs are inferred via semantic analysis. In linguistics, current approaches for semantic analysis include Componential Analysis (CA) [13], Natural Semantic Metalanguage (NSM) [14], and Lexical Conceptual Structure (LCS) [12], etc. In the CA framework, the results and tools for C&B actions are enumerated and analyzed: e.g., Wan proposed the semantic primitives including tool, direction, force, contact, object, result, and intention, etc. [15]. In the NSM framework, C&B verbs are analyzed with a six-part semantic template, including motivational scenario, instrument, usage pattern of instrument, result on the object, and the potential outcome [16]. In the LCS framework, the C&B verbs are classified into cutting-type and breaking type, and associated with distinct argument structure and syntactic privileges [17].

With the implemented semantic analysis methods (i.e., CA, NSM, and LCS), motion primitives of the C&B verbs can be explicitly interpreted. However, as the analysis methods are primarily based on qualitative, expert, and empirical knowledge, it is unable to quantitatively interpret the C&B verbs into the trajectory patterns of the robot end-effector. Moreover, it is also difficult to validate the interpreted meaning, or to formulate unified criteria among researchers.

To address the difficulties in interpreting the C&B verbs into robot actions, we propose a mechanics-based approach. The rationale of the approach is that the semantics of C&B verbs (as well as its superset --- action verbs) are rooted in human’s body experience. This body experience corresponds to the physical interaction, where mechanics is the underlying principle, between the world and human’s motor/sensor. Accordingly, the trajectory pattern of the actions can be extracted from the instructed verbs from the mechanics perspective.

The rest of this paper is organized as follows: In Section 2, we propose the representations of C&B verbs based on mechanics. In Section 3, applications of the proposed representations in trajectory generation is demonstrated. Conclusion is given in Section 4.

2. Contact, Twist, and Wrench Representations of C&B Verbs
Completeness and minimalism are the crucial requirements for interpreting the physical meaning of C&B verbs: If the describing parameters are insufficient (violating completeness), the meaning cannot be wholly expressed; if the parameters are redundant (violating minimalism), it implies that we had either introduced an irrelevant parameter, or had taken multiple “facets” of a single parameter into the description.

Cutting involves the relative motion and the interacting force between the tool (including human hands) and the object. On the one hand, without motion or force, “cutting” cannot be realized (necessity). On the other hand, when the motion and force have been specified, the cutting action is uniquely defined (sufficiency).

There are infinite number of choices to represent motion and force, but mechanicians pursue a “complete and minimal” representation, i.e., the complete description of motion and force with the minimal number of parameters. After decades of investigations, the twist and wrench [18] prevails in representing motion/force, for its completeness and minimalism, which exactly match the requirement of physically interpreting a verb. Therefore, the following framework for representing C&B verbs is also based on twists and wrenches. Besides, contact is also the precondition for the cutting action (Historically, it was not a common practice to remotely cut/break an object without contact.). Next, we discuss the role of contact.

2.1. Contact between Tool and Object
Contacts may differ in many aspects, but the radical difference is the dimension of the contact region, namely (illustrated in Fig. 1):
Figure 1. Dimensions of contact regions.

- 0-dimensional point contact (Microscopically, all contacts are point contacts. But this fact is beyond human’s sensors’ resolutions.), e.g., 剁 (cì);
- 1-dimensional line contact, e.g., 切 (qiē);
- 2-dimensional area contact, e.g., 砸 (zá);

2.2. Twist Vector

Twist is the core of motion description, which is under a defined coordinate frame (hereafter referred to as frame for brevity). For C&B verbs that involve a tool and an object, two frames naturally exist: the object frame and the tool frame, which are fixed to the object and the tool, respectively (as shown in Fig. 2). The frame definitions and notations are given in Fig. 2.

Figure 2. Frame definitions.

The object frame is denoted by $\Sigma_o$, and the tool frame is denoted by $\Sigma_t$.

The origin of $\Sigma_o$ is a point in the contact region, its $z_o$ axis is the surface normal at the origin, $x_o$ is along the contact line (for line contact) or arbitrary (for point and area contacts), and $y_o$ makes a right-hand orthogonal frame with $x_o$ and $z_o$.

The origin of $\Sigma_t$ is an arbitrary point on the tool. $x_t$ is along the contact line (for line contact) or arbitrary, $y_t$ is along the normal of the expected separation plane (for line contact) or arbitrary, and $z_t$ makes a right-hand orthogonal frame with $x_t$ and $y_t$.

Given the frames $\Sigma_o$ and $\Sigma_t$, the twists can be defined: The twist of object $a$ under frame $b$ is defined as:

$$b s_a = [\omega, v] = [\omega_x, \omega_y, \omega_z, v_x, v_y, v_z],$$

Where $\omega = [\omega_x, \omega_y, \omega_z]$ are object $a$’s angular velocities, and $v = [v_x, v_y, v_z]$ are object $a$’s translational velocities. Besides, as the twist varies with time, we use $b s_a(t), \omega(t),$ and $v(t)$ to show their dependency on time, and let $t = 0$ to be the start moment of contact.

Objects’ motions can be completely and minimally represented by twists [18], with which the C&B verbs can be interpreted upon its direction and magnitude. For example:

For 切 (qiē), the tool’s direction in the object frame $\Sigma_o$ is along $-z_o$; the object’s direction in the tool frame $\Sigma_t$ is along $-z_t$. Note that, instead of an exact trajectory, a verb corresponds to a family of trajectories that satisfy a set of motion constraints, for 切, the constraints are

$$^o s_t \cdot [0,0,0,0,-1] \geq \alpha,$$

$$^t s_o \cdot [0,0,0,0,-1] \geq \alpha,$$
Where $\alpha > 0$ is a user-specified threshold (e.g., 0.7), which can be obtained from survey or learning from experiment data. Similarly, other C&B verbs can also be quantitatively described in the defined coordinate frames, e.g.:

- For 切 (qiē), the tool’s direction in the object frame $\Sigma_o$ is along $-z_o$; the object’s direction in the tool frame $\Sigma_t$ is along $-z_t$;
- For 钻 (zuàn), the tool’s rotation axis in $\Sigma_o$ is parallel to $z_o$.
- For 割 (gē), the tool’s direction in $\Sigma_o$ is parallel to $x_o$; the object’s direction in $\Sigma_t$ is parallel to $x_t$;
- For 削 (xiāo), the tool’s direction in $\Sigma_o$ is parallel to $y_o$; the object’s direction in $\Sigma_t$ is along $y_t$;
- For 刮 (guā), its tool’s direction in $\Sigma_o$ is identical to 削 (xiāo), but the object’s direction in $\Sigma_t$ is parallel to $y_t$;
- For 砍 (kǎn), its direction is identical to 切 (qiē), but its magnitude at $t = 0$ is significantly larger.

The above descriptions are summarized and quantified in Table 1.

2.3. Wrench Vector
Object may break under pure force/torque, even without the motion of the tool. The force/torque is described by the wrench vector. In frame b, the force/torque acting on object a is defined as

$$\mathbf{w}_b = [f, \tau] = [f_x, f_y, f_z, \tau_x, \tau_y, \tau_z],$$

Where $f = [f_x, f_y, f_z]$ denote the forces along the axes, and $\tau = [\tau_x, \tau_y, \tau_z]$ are the torques around the axes. In Mandarin Chinese, the motionless C&B verbs can be interpreted depending on the magnitudes of the force and the torque. For example:

- For 揪 (sī), etc., the magnitude of $f$ is significantly larger than that of $\tau$, as shown in Fig. 3a.
- For 破 (bāi), etc., the magnitude of $\tau$ is significantly larger than that of $f$, as shown in Fig. 3b.

![Figure 3. Magnitudes of wrenches.](image)

The above descriptions are summarized and quantified in Table 1.

By now we have arrived at the representations of C&B verbs using contact dimensions, twists, and wrenches. The proposed formulation is based on classic mechanics and easy to be quantitatively verified. Next, we briefly introduce the applications of the proposed representations.

3. Applications in Motion Generation
Since the proposed descriptions in Section II is formal and quantitative, it is straightforward to apply those descriptions on cut-motion generations. For example, Fig. 4 shows the generated motion corresponding to 切 (qiē), Fig. 5 gives the motion corresponding to 削 (xiāo), and Fig. 6 gives the motion corresponding to 割 (gē).
Figure 4. Generated motion sequence of 切 (qiē).

Figure 5. Generated motion sequence of 削 (xiāo).

Figure 6. Generated motion sequence of 割 (gē).

Note that verbs’ meanings are fuzzy: Instead of representing a unique twist or wrench, the meaning of a C&B verb actually corresponds to a group of motions that satisfy a set of constraints (rightmost column of Table 1).

Table 1. Descriptions and representations of the example verbs.

| Verb | Dim | Description | Twist and wrench representation |
|------|-----|-------------|--------------------------------|
| 刺 | 0   | Tool’s velocity direction is perpendicular to the object’s surface and points inwards. | $-\mathbf{v}_t \cdot \mathbf{z}_o > \alpha \|\mathbf{v}_t\|$. |
| 切 | 1   | Same as (戳), with initial velocity (at the moment of contact) close to zero. | Same as (戳), with $\|\mathbf{v}_t(0)\| \leq \beta$ |
| 割 | 1   | Tool’s velocity direction is parallel to the line of contact. | $\mathbf{v}_t \cdot \mathbf{x}_o > \alpha \|\mathbf{v}_t\|$, $\mathbf{v}_o \cdot \mathbf{z}_t > \alpha \|\mathbf{v}_o\|$. |
| 削 | 1   | Tool’s velocity direction is parallel to $\mathbf{y}_o$, object’s velocity direction is parallel to $\mathbf{z}_t$. | $\mathbf{v}_t \cdot \mathbf{y}_o > \alpha \|\mathbf{v}_t\|$, $\mathbf{v}_o \cdot \mathbf{z}_t > \alpha \|\mathbf{v}_o\|$. |
| 割 | 1   | Same as (削), but the object’s velocity direction is parallel to $\mathbf{y}_t$. | $\mathbf{v}_t \cdot \mathbf{y}_o > \alpha \|\mathbf{v}_t\|$, $\mathbf{v}_o \cdot \mathbf{z}_t > \alpha \|\mathbf{v}_o\|$. |
| 砍 | 1   | Same as (切), but the magnitude of initial velocity is significant. | Same as (切), but $\|\mathbf{v}_t(0)\| > \beta$ |
| 钻 | 0   | The tool’s angular velocity is perpendicular to the object’s surface. | $\|\mathbf{w}_t \cdot \mathbf{z}_o \| > \alpha \|\mathbf{w}_t\|$. |
| 砍 | 2   | Same as (砍) | Same as (砍) |
| 撕 | 0,1,2 | The force magnitude is significantly larger than that of the torque. | $\|\mathbf{f}\| > \gamma \|\mathbf{r}\|$. |
| 撕 | 0,1,2 | The torque magnitude is significantly larger than that of the force. | $\|\mathbf{f}\| > \gamma \|\mathbf{r}\|$. |

Note that $\alpha$, $\beta$, and $\gamma$ are user-specified thresholds, which can be obtained from surveys or learned from experiment data.

Consequently, the way of generating motion is application-specific: For visualization of verb meanings, motions are randomly generated within the constraints (which are also the cases of Fig. 4,
Fig. 5, and Fig. 6). For robot manipulation, the final motion is sought according to a given optimality criterion, while our proposed representations serve as constraints that largely reduce the search space and search time. However, details of the constrained optimization are beyond the scope of this paper.

4. Conclusion
Using contact dimension, twist, and wrench, we propose a mechanics-based approach of interpreting C&B verbs into trajectory patterns, and formulate a set of constraints on those patterns as their quantitative measures of actions. Since the description of motion/force involved in the cutting action is rigid, complete and minimal, the interpreted verb meanings are formal, unique, and computable. Therefore, the proposed method can be extended to both natural language processing and robot manipulation: For natural language processing, the formal and quantitative descriptions of verb meanings facilitate the investigations in categorization, syntax-semantics interface, etc. For robot manipulation, the interpreted verb meanings correspond to the humanlike strategy in planning and motion generation.

However, the presented formulation only covers the Mandarin Chinese C&B verbs, while the larger set of manipulation verbs in more languages are still to be investigated. These issues are to be addressed in our future work.

5. Acknowledgement
This work is supported by the National Social Science Fund of China (NSSFC) under grant number 18CYY056.

6. References
[1] Winograd T 1971 Procedures as a representation for data in a computer program for understanding natural language Technical report MIT
[2] Thomas K, Stefanie T, Deb R, and Nicholas R 2010 Toward understanding natural language directions. International Conference on Human-Robot Interaction pp 259 – 266
[3] Dzifcak J, Scheutz M, Baral C, and Schermerhorn P 2009 What to do and how to do it: Translating natural language directives into temporal and dynamic logic representation for goal management and action execution Proceedings of 2009 IEEE International Conference on Robotics and Automation
[4] Hao D and Allen P 2012 Semantic grasping: Planning robotic grasps functionally suitable for an object manipulation task, Proceedings of the 2012 IEEE International Conference on Intelligent Robots and Systems pp 1311–17
[5] Yang Y, Yiannis A, Cornelia F, and Eren E 2015 Learning the semantics of manipulation action Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics Conference pp 676–86
[6] Hale K and Keyser S 1987 A view from the middle Lexicon Project Working Papers 10. Center for Cognitive Science (Cambridge, Massachusetts: MIT)
[7] Toth N and Schick K 1993 Early stone industries and inferences regarding language and cognition. In Gibson K R and Ingold T (eds.), Tools, Language and Cognition in Human Evolution 346–362(Cambridge. Cambridge University Press) pp 346–362
[8] Majid A, Bowerman M, Van Staden M and Boster JS 2007 The semantic categories of cutting and breaking events: A crosslinguistic perspective Cognitive Linguistics 18(2) pp 133–52
[9] Majid A, Boster J and Bowerman M 2008 The cross-linguistic categorization of everyday events: A study of cutting and breaking. Cognition 109 pp 235–50
[10] Wang C, Hong C and Saji N 2018 Hanri qiege lei dongci fanchou hua de kua yuyan shizheng yanjiu (An empirical study on verbs of cutting and breaking in Chinese and Japanese). Foreign Language Teaching and Research 50(4) pp 516–28
[11] Levin B 1993 English Verb Classes and Alternations (Chicago: University of Chicago Press)
[12] Levin B and Rappaport Hovav M 2011 Lexical Conceptual Structure Maienborn C Von Heusinger K and Portner P (eds.) Semantics: An International Handbook of Natural Language Meaning. (Berlin: De Gruyter Mouton) pp 420–40
[13] Leech G 1981 Semantics: The Study of Meaning (UK: Penguin)
[14] Wierzbicka A 1972 Semantic Primitives (Frankfurt: Athenäum)
[15] Wan Y 2018 Hanyu qiege lei dongci cihua moshi fenxi (Analysis on the lexicalization patterns of Mandarin Chinese cut verbs) Linguistic Research 146(1) pp 1–10
[16] Goddard C and Wierzbicka A 2009 Contrastive semantics of physical activity verbs: cutting and chopping in English, Polish, and Japanese Language Sciences 31 pp 60–96
[17] Levin B and Rappaport Hovav M 2013 Lexicalized meaning and manner/result complementarity in Arsenijević B, Gehrke B, and Marin R (eds.) Subatomic Semantics of Event Predicates (Dordrecht : Springer) pp 49–70
[18] Selig J 2000 Geometric Fundamentals of Robotics (Singapore: World Scientific Publishing)