

The grammar of dish names in Mandarin Chinese

Juan Zhong and Jean-Pierre Koenig

University at Buffalo

juanzhon@buffalo.edu jpkoenig@buffalo.edu

1 Introduction

HPSG assumes that linguistic generalizations, phrasal as well as lexical, come in all sizes: some more general and cross-linguistically attested than others. The existence of specialized-domain grammars, that is, grammatical patterns that are specific to a particular semantic domain, support this approach to grammar design, as HPSG lexical and phrasal sort hierarchies are particularly well suited to model specialized-domain grammars (see Koenig & Michelson 2021 for the grammar of possession in Oneida, as well as Gross 1995 for the grammar of emotions in French, and Smith 1999 for the grammar of number names in English). In this paper, we expand the range of semantic domains for which a specialized-domain grammar might be needed by considering the grammar of dish names in Mandarin.

There are three broad aspects of Mandarin Chinese dish names (MDNs) of particular interest to this paper. First, MDNs can consist of long sequences of words without any overt evidence of structure. For example, (8) simply consists of a list of seven content words. In addition to the various ingredients (i.e., black pepper, garlic, beef, and cooked rice) and the cooking method (i.e., stir-fry), information about the shape and size of the ingredient beef (i.e., diced) can be expressed in an MDN. Second, structural ambiguity can arise when a verb is followed by a noun, as that sequence can be interpreted either as a dish name (9) or as a verb phrase (10) depending on the context, a structural ambiguity that is predicted if MDNs constitute a special kind of expressions (a specialized constructional sort). Third, and a central point of our paper, the grammar of MDNs differs from the rest of the grammar of Mandarin. For example, although some MDNs can include a substring of the form NVN, that substring does not follow standard Mandarin clausal syntax, as the first noun is not the subject selected by the following verb, as is usually the case in an SVO Mandarin sentence. For example in (11), the egg is not the proto-agent but the proto-patient of the following verb. We show in the paper that the grammar of Mandarin dish names does not pattern like standard NN or VN compounds either (Packard, 2000). In other words, dish names have a *sui generis* grammar, irreducible to more standard Mandarin constructions.

MDNs have been previously treated as complex compound words (Chao, 1968) and recent studies have explored the structure of a small subset of types of MDNs using Construction Grammar (Song, 2022) or lexical semantics (Shen & Li, 2024). Although previous studies offer descriptive insights, they do not cover the entire range of possible dish names, particularly the ordering constraints for dish names that contain three or more words, nor do they provide a detailed formal analysis of MDNs as our paper does. Overall, we argue that MDNs are best modeled as a sequences of lists of expressions denoting semantically-defined categories (Dish Categories, Ingredients, Methods, Properties) with several constraints on each of these sublists. We further show that HPSG’s extensive use of *lists* to encode constituency is particularly well-suited to model the grammar of MDNs.

2 Dataset and descriptive generalizations

Our corpus of MDNs is the Mandarin Dish Name Dataset, a list of 1,996 Chinese dish names published in *Zhongwen Caidan Yingwen Yifa* (“The Way of Translating Chinese Menus into English”) by the Beijing Foreign Affairs Office in 2007. The list was created to standardize English translations for the Beijing 2008 Olympic Games and provides index numbers, simplified Chinese characters, and English translations for each dish name. A duplicate entry and 19 semantically-opaque entries (such as (15)) were excluded, leaving 1,976 dish names for analysis. The expressions making up each dish name were annotated for lexical categories (i.e., noun, verb, adjective, and classifier) and semantic categories (i.e., ingredient, method, property, and dish category¹). Overall, MDNs consist of five sublists of expressions as shown in (1)². Only the sequences of sublists enumerated in (2) occur in our dataset (and are possible, according to an informal survey of additional dish names); see (12) for an example of pattern (2f) and of our annotation of corpus examples. Note that properties are predicated of the entire dish name and that cooking methods (e.g., stir-frying) applies to all members of the Ingredients_A and Ingredients_B lists.

(1) list(Prop(erties)) $\oplus$ list(Ingr(edients)_A) $\oplus$ list(Methods) $\oplus$ list(Ingr(edients)_B) $\oplus$ list(Cat(egory))

- | | |
|--|--|
| <p>(2) a. Prop$\oplus$Ingr_B
 b. Ingr_A$\oplus$Ingr_B
 c. Methods$\oplus$Ingr_B
 d. Prop$\oplus$Ingr_A$\oplus$Ingr_B
 e. Prop$\oplus$Methods$\oplus$Ingr_B
 f. Ingr_A$\oplus$Methods$\oplus$Ingr_B
 g. Prop$\oplus$Ingr_A$\oplus$Methods$\oplus$Ingr_B</p> | <p>h. Prop$\oplus$Cat
 i. Ingr_B$\oplus$Cat
 j. Prop$\oplus$Ingr_B$\oplus$Cat
 k. Ingr_A$\oplus$Ingr_B$\oplus$Cat
 l. Methods$\oplus$Ingr_B$\oplus$Cat
 m. Prop$\oplus$Methods$\oplus$Ingr_B$\oplus$Cat
 n. Prop$\oplus$Ingr_A$\oplus$Ingr_B$\oplus$Cat
 o. Ingr_A$\oplus$Methods$\oplus$Ingr_B$\oplus$Cat
 p. Prop$\oplus$Ingr_A$\oplus$Methods$\oplus$Ingr_B$\oplus$Cat</p> |
|--|--|

3 An HPSG analysis of MDNs

To account for the possible sequences enumerated in (2), we propose the MDN construction in (3) as well as the constraints in (4)-(7). As our description of the construction suggests, MDNs are linguistic signs that consist of the concatenation of five distinct, *possibly empty*, lists of signs. More specifically, the construction stipulates that a dish name consists of a list of properties (extrinsic, e.g. place names, personal names, or intrinsic, e.g. flavor, texture, in that order), a list of (minor) ingredients, a list of cooking or serving methods, a list of (major) ingredients, and finally a list of dish categories.

¹By dish category we mean labels such as *soup* in the dish name *onion soup*.

²In accordance with the distinction between the concept of *zhǔliào* ‘main ingredient’ and *pèiliào* ‘accompanying ingredient’ in Chinese culinary traditions Xiao (1992), we categorize ingredients into two lists: Ingredients_A for minor ingredients and Ingredients_B for major ingredients.

(3) The Mandarin Dish Name Construction:

$$\begin{array}{l}
 \text{dish-name-constr} \\
 \text{PHON } \text{list} \\
 \text{SYNSEM } [\text{HEAD } \boxed{6}] \\
 \text{DTRS } \boxed{0} \left(\begin{array}{l} \oplus \boxed{1} \text{list} \left(\begin{array}{l} \text{PHON } \text{list} \\ \text{POS } \text{noun} \vee \text{adj} \\ \text{CONT } \text{property} \end{array} \right) \\ \oplus \boxed{2} \text{list} \left(\begin{array}{l} \text{PHON } \text{list} \\ \text{POS } \text{noun} \\ \text{CONT } \text{minor_ingredient} \end{array} \right) \oplus \boxed{3} \text{list} \left(\begin{array}{l} \text{PHON } \text{list} \\ \text{POS } \text{verb} \\ \text{CONT } \text{method} \end{array} \right) \\ \oplus \boxed{a} \left(\begin{array}{l} \boxed{4} \text{list} \left(\begin{array}{l} \text{PHON } \text{list} \\ \text{POS } \text{noun} \\ \text{CONT } \text{major_ingredient} \end{array} \right) \oplus \boxed{5} \text{list} \left(\begin{array}{l} \text{PHON } \text{list} \\ \text{POS } \text{noun} \\ \text{CONT } \text{category} \end{array} \right) \end{array} \right) \end{array} \right) \\
 \text{CONT } [\text{MAIN } \text{dish}]
 \end{array}
 \right)$$

A crucial aspect of our analysis is that there is no internal constituency among sublists $\boxed{1}$ through $\boxed{5}$ in (3). Two facts make the possibility of referring to all sublists of daughters—which the flat structure we propose affords us—particularly useful. As none of the sublists are required, to ensure the right order between five (possibly empty) sublists all of which can be empty, positing a hierarchical structure would either require totally empty categories (i.e., phonologically, morphosyntactically, and semantically empty categories) or a multiplication of subsorts of the MDN construction. Similarly, a hierarchical structure among the five sublists in (3) would make it harder to model the fact that the last daughter must be a member of the last two lists despite the fact that neither list is required.

There are several constraints that limit the range of possible MDNs. (4)-(7) provide both English and HPSG descriptions of all constraints on MDNs we have identified. The statement of the constraints is simplified and we are liberal in our use of names of relations for expository purposes; tags correspond to the tags in the description of the construction in (3).

- (4) a. Minimum length: A dish name must contain at least two daughters.
b. $\text{length}(\boxed{0}) \geq 2$
- (5) a. Cardinality restrictions: Lists $\boxed{3}$ and $\boxed{5}$ contain at most one member.
b. $\text{length}(\boxed{3}) \leq 1 \wedge \text{length}(\boxed{5}) \leq 1$
- (6) a. Final daughter restriction: The last daughter must belong to List $\boxed{4}$ or List $\boxed{5}$.
b. $\left[\text{DTRS } \text{list} \oplus \langle \boxed{7} \rangle \right] \Rightarrow \text{member}(\boxed{7}, \boxed{a})$
- (7) a. Head daughter selection: The last daughter is the syntactic head of the MDN.

$$\text{b. } \text{dish-name-constr} \Rightarrow \begin{array}{l} \left[\begin{array}{l} \text{HEAD } \boxed{6} \\ \text{CONT } [\text{DR } \boxed{8}] \\ \text{DTRS } \text{list} \left([\text{CONT|DR } \text{none}] \right) \oplus \left\langle \begin{array}{l} \text{HEAD } \boxed{6} \\ \text{CONT|DR } \boxed{8} \text{index} \end{array} \right\rangle \end{array} \right]
 \end{array}$$

Semantically, we assume that only the last daughter introduces a discourse referent (in the DRT sense, Kamp & Reyle 1993). Other daughters include (free) variables that can serve as arguments of predicates, but not discourse referents, along the line of the analysis of French indefinite *on* in Koenig (1999) or Hungarian bare object NPs in Farkas & de Swart (2003). The test we use to determine which daughter in an MDN introduces a discourse referent is illustrated in (13). When quantifying over individual dishes described by a dish name, the dish name can be shortened and only include the dish category, if there is one, or an

ingredient from the Ingredients_B list, if not, as seen in (13), but no other daughter. In other words, the last daughter can stand in for the entire dish and provide the discourse referent that the numeral-classifier phrase quantifies over (we omit discussion of difficult cases in the abstract).³

Although there is no evidence of hierarchical structure between the five component lists in the *dish-name-constr*, members of individual lists can themselves consist of lists of daughters. Properties members may take the form A-A (e.g., *suān-tián* ‘sour-sweet’); Ingredient_A/_B members can take the form A-N (e.g., *xiān-qíézi* ‘fresh eggplant’), N-N (e.g., *Dōngpō-ròu* ‘Dongpo pork’), V-N (e.g., *hóngshāo-ròu* ‘red-braised pork’), or N-CL (e.g., *dòufu-sī* ‘tofu shred’); Methods members can take the form N-V (e.g., *tàn-shāo* ‘char-grill’), but also A-V (e.g., *gān-biān* ‘dry-stir.fry.in.hot.oil’) or V-V (e.g., *jiān-niàng* ‘pan.fry-stuff’). Additionally, structurally complex ingredients (or methods) can themselves be the output of a sort-shifting MDN to ingredients (or methods) unary-branching construction. The MDN (14) *hóng-shāo niú-tóu kòu liáoshēn* ‘red-braise ox-head button sea.cucumber’, for example, contains the ingredient *hóng-shāo niú-tóu* ‘red-braise ox-head’ which itself is a possible dish name.⁴

4 General discussion

MDNs add to the list of languages (English, French, and Oneida) and semantic domains (number names, expressions of emotions, possession) with a distinct local grammar, to borrow Gross (1995)’s term. Like the expression of possession in Oneida according to Koenig & Michelson (2021), MDNs build on morphosyntactic tools made available by the rest of the grammar of Mandarin, in particular the wide availability of multi-word compounding, but alter those tools in unpredictable ways to fit the particular properties of the concepts that belong to the semantic domain of dish names. More specifically, although Mandarin makes extensive use of compounding, the sequence of words in MDNs differs from ordinary compounding in critical ways. Consider, for example, the lists of ingredients in patterns (2b), (2d), (2k), and (2n). These sequences of nouns do not fit standard Mandarin NN compounds. First, there can be more than two ingredients across the two lists; in fact, Ingredients_A on its own can consist of more than two members. Second, the meaning of the entire list of ingredients does not fit any of the possible semantic correlates of compounding discussed in Packard (2000). Basically, the semantic contribution of both Ingredients_A and Ingredients_B lists is that the dish contains all members of both lists. Mandarin speakers seem to have adapted a common strategy in their language (compounding) to create new object categories to fit the demands of the semantic domain of dish names: the list of ingredients can be longer than two and the semantic contribution of the sequences of ingredients is not to relate ingredients to each other (as for typical NN compounds) but to the overall dish they are part of and that the MDN names.

At a theoretical level, MDNs tell us something about the architecture of grammars, namely that languages can include idiosyncratic, but regular patterns for semantically coherent domains, and that theories of natural language should provide the tools to model instances of such “bricolage” to borrow a term from Koenig & Michelson (2021). But the grammar of MDNs provide us with two additional lessons. The first is “technological”: treating daughters as (possibly empty) lists of expressions is a very useful tool when modeling templatic syntax (Good, 2015), particularly, complex patterns such as those in (2). The second is that *kinds* in the sense of Carlson (1977) can be associated not just with lexical categories (nouns in particular); they can also be associated with complex expressions such as MDNs. McNally & Boleda (2004) propose that nouns include an implicit kind argument and that relational adjectives are predicated

³More precisely, the dish name describes a unique kind and the discourse referent of the last daughter realizes this kind in Carlson’s (1977) sense. We omit this complication here for reasons of space and will discuss it more fully in the paper.

⁴It is tempting to treat the sequence Methods⊕Ingredients_B as a subconstituent in light of the fact that cooking methods must be followed by Ingredients_B. Note, though, that when Ingredients_A also precedes the cooking method verb(s), they must be within the scope of the same verb(s), and that the syntactic head, if no dish category is present, is a member from Ingredients_B. Both facts militate against this subconstituent analysis.

of that argument. We propose that each MDN, despite being a complex morphosyntactic expression, also includes an implicit kind argument, a *kind* of dish just as simple nouns can. This proposal accounts for the fact that MDNs are semi-compositional: each daughter contributes to the meaning of the entire MDN (we will detail how for each of the four categories listed in (1)), but the meaning of the whole does not reduce to a set of ingredients, properties, methods, and dish categories. There is an individuality to the dish named by the MDN that goes beyond knowing what is in it or how you cook it. This is, we claim, because each MDN is predicated of a unique kind. The fact that MDNs denote a kind, just as words do, probably explains why Chao (1968) suggests they are compounds despite their complexity unlike any other kind of compounds in Mandarin and their syntax-like properties (individual daughters of a *dish-name-constr* construct can be morphosyntactically complex or even derived from another dish name).

Examples

- (8) hēijiāo xiāng suàn niúliǔ lì chǎo fàn (705)
 black.pepper fragrant gallic beef.tenderloin grain stir.fry cooked.rice
 N [A N]_N [N CL]_N V N
 ‘stir-fried cooked rice with diced beef tenderloin, black pepper and garlic’
- (9) wǒ yào zhá jī chì (constructed)
 I want deep.fry chicken wing
 S V [V [N N]_N]_N
 ‘I want fried chicken wings.’
- (10) wǒ yào zhá jī chì. (constructed)
 I want deep.fry chicken wing
 S V V [N N]_N
 ‘I want to deep-fry chicken wings.’
- (11) dàn chǎo fàn (Chao 1968: 505)
 egg stir.fry cooked.rice
 N V N
 ‘stir-fried cooked.rice with eggs’
- (12) Ingr_A⊕Methods⊕Ingr_B (0325)
 mùguā dùn chì
 papaya stew shark’s.fin
 N V N
 INGREDIENT METHOD INGREDIENT
 ‘shark’s fin soup with papaya’
- (13) a. yī pán suān-tián pào-cài (0812)
 one plate sour-sweet pickle-vegetable
 Num CL [A-A]_A N
 ‘a plate of sweet and sour pickled vegetables’
 b. yī pán pào-cài
 one plate pickle-vegetable
 Num CL N
 ‘a plate of pickled vegetables’

- (14) **hóng-shāo niú-tóu kòu liáoshēn** (373)
red-braise ox-head button sea.cucumber
 [V N]_N V N
 ‘braised sea cucumber with ox head’
- (15) **fó tiào qiáng** (1635)
buddha jump wall
 N V N
 ‘Fotiaoqiang (Steamed Abalone with Shark’s Fin and Fish Maw in Broth)’

References

- Carlson, Greg. 1977. A unified analysis of English bare plurals. *Linguistics and Philosophy* 1. 413–457.
- Chao, Yuen Ren. 1968. *A grammar of spoken chinese*. University of California Press.
- Farkas, Donka & Henriëtte de Swart. 2003. *The semantics of incorporation* Stanford Monographs in Linguistics. Stanford, CA: CSLI Publications.
- Good, Jeff. 2015. *The Linguistic Typology of Templates*. Cambridge University Press 1st edn. doi:10.1017/CBO9781139057479. <https://www.cambridge.org/core/product/identifier/9781139057479/type/book>.
- Gross, Maurice. 1995. Une grammaire locale de l’expression des sentiments. *Langue française* 105. 70–87.
- Kamp, Hans & Uwe Reyle. 1993. *From discourse to logic* (Studies in Linguistics and Philosophy 42). Dordrecht: Kluwer.
- Koenig, Jean-Pierre. 1999. French body-parts and the semantics of binding. *Natural Language and Linguistic Theory* 17(2). 219–265.
- Koenig, Jean-Pierre & Karin Michelson. 2021. Specialized-domain grammars and the architecture of grammars: Possession in Oneida. *Journal of Linguistics* 57(3). 635–674.
- McNally, Louise & Gemma Boleda. 2004. Relational adjectives as properties of kinds. In O. Bonami & P. Cabredo Hofherr (eds.), *Empirical issues in formal syntax and semantics* 5, 179–196. Paris: Centre National de la Recherche Scientifique.
- Packard, Jerome L. 2000. *The morphology of chinese: A linguistic and cognitive approach*. Cambridge University Press.
- Shen, Yuqi & Jiapan Li. 2024. The formation features and structural mechanism of modern chinese “vcooking+n” structure. In Minghui Dong, Jia-Fei Hong, Jingxia Lin & Peng Jin (eds.), *Chinese lexical semantics*, vol. 14514, 116–129. Springer Nature Singapore. doi:10.1007/978-981-97-0583-2_10. https://link.springer.com/10.1007/978-981-97-0583-2_10. Series Title: Lecture Notes in Computer Science.
- Smith, Jeffrey. 1999. English number names in HPSG. In Gert Webelhuth, Jean-Pierre Koenig & Andreas Kathol (eds.), *Lexical and constructional aspects of linguistic explanation* Studies in Constraint-Based Lexicalism, 145–160. Stanford: CSLI Publications.
- Song, Zuoyan. 2022. The characteristics of Construction Morphology and its enlightenment on the study of Chinese word formation: The case of the naming patterns of Chinese dishes. *Language Teaching and Linguistic Studies* (2). 88–102.
- Xiao, Fan (ed.). 1992. *Zhongguo peng ren ci dian*. Peking] : [Shanghai: Zhongguo shang ye chu ban she ; Xin hua shu dian Shanghai fa xing suo jing xiao di 1 ban edn.