Collectives, Singulatives, Mass and Count:

A subpredicative approach to countability informed by the collective/singulative systems of Welsh and Arabic

Victoria Noble

Doctor of Philosophy

University of York

Language and Linguistic Science

March 2025

Abstract

In this thesis, I investigate the collective/singulative systems of Welsh and Arabic which are shown to have the same descriptive behaviours as the English object mass system and the Mandarin Chinese general number system. Explicitly, object mass, collective and general number are shown to be number neutral and uncountable, and the singulative marker is classificatory in nature. Based on this, I propose an analysis where object mass, collective and general number systems share the same underlying primitive features. Following this, I also account for the differences between these systems, including: the role of construed aggregation in the Welsh, Arabic and English object mass/collective systems that is absent in the Mandarin Chinese general number system; why the singulative marker in Welsh cannot pluralise, but the singulative marker in Arabic may do so; and the degree of nominal flexibility (or lack thereof) of collective/singulative terms in 'universal grinder' contexts.

Collective/singulative languages offer a curious paradigm for meaning/form correspondences: alongside a singular/plural system, they experience a minor collective/singulative system where the singular form is morphologically marked. There are two prominent theoretical approaches for collective/singulative systems which have complementary strengths and limitations. Grimm (2012b, 2018)'s approach prioritises incorporating perceptual influence of construed aggregation at the lexical level to account for the fact that the collective/singulative class typically refers to individuals that come in groups or aggregations. However, it is not clear how this system captures grammatical similarities between collective/singulative and

general number systems, which are not associated with aggregation. On the other hand, Mathieu (2012a)'s (*et. seq.*) approach prioritises crosslinguistic parsimony in the syntax, where the function of singulatives and classifiers are subsumed under a single syntactic head. However, it is not clear how this system captures how or why aggregation influences collective/singulative but not general number systems.

The analysis I offer unites the strengths of Grimm and Mathieu's approaches in the style of de Vries and Tsoulas (2021)'s `SUBPREDICATIVE ICEBERG SEMANTICS`. I expand Subpredicative Iceberg Semantics by incorporating a Grimm-inspired aggregation semantics to account for collective/singulative distribution as a minor number system, but also propose functions of number that unite object mass, collective and general number under the number neutral uncountable umbrella, while treating classifier and singulative semantics as akin. I offer a split analysis of the plural of the singulative where number features are distributed across the nominal spine in a Mathieu inspired way. The semantics of the account is designed such that the plural of the singulative is attested in Arabic, but is ruled out under economy in Welsh.

Finally, this thesis goes one step further than previous explorations of collective/singulative systems: I explore the use of collective/singulative terms in 'universal grinder' contexts (see Pelletier, 1975; Borer, 2005a; Cheng et al., 2008; Rothstein, 2017). I show that grinding effects for collective terms in Welsh and Arabic loosely parallel what is seen for Mandarin Chinese general number terms, however the grinding effects for singulative terms is not uniform across Welsh or Arabic (Welsh singulatives can be 'ground', Arabic singulatives cannot). The data questions whether there exists a single 'universal grinder' within and across languages, and subsequently I propose a multi-source approach to grinder readings.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Acknowledgements

In this thesis, I talk a lot about mereological modelling. It is only fitting, then, that I properly acknowledge those who have been *part* of my PhD journey.

First and foremost, I would like to thank my PhD supervisor, George Tsoulas. I have known George since my undergraduate years at York, and he has no doubt been the biggest influence on my linguistic journey. Every conversation I have had with George has been a delight. From intellectually stimulating linguistic business to the more personal, George has been a fantastic mentor. I feel honoured to have worked with him for so long. Thank you, George, for all the time and energy you have invested in me throughout the years.

There are other members of the semantics/syntax group at York who have variously inspired and shaped my thoughts and experiences throughout this journey. First, Norman Yeo. Norman was my first ever tutor in semantics and syntax. Looking back to how much fresh 18 year me old loved your seminars, I think it was inevitable that I would end up doing a PhD in semantics. Thank you for inspiring me back then, and continuing to support me over the years. If you think about it, me doing this thesis is entirely your fault - and I couldn't be more grateful. Next, Nino Grillo. Nino is an all round fantastic human. He has never failed to make me smile, and has always encouraged me to be confident in myself and my ideas. Toward the end of my PhD, Nino's support meant the difference between finishing this PhD or not. I am so very grateful for your never ending support and encouragement. Next, Eytan Zweig. I was incredibly lucky enough to be Eytan's teaching assistant. I was so happy to teach on what was my favourite module as an undergraduate. Thank you Eytan, for giving me that opportunity. I really enjoyed working with you. Next, Pino Longobardi. Pino has been an excellent source of confidence throughout this journey. From my masters degree onwards, Pino has never failed to believe in me or my ideas; he has always made me feel more confident in everything I have proposed, even if it turned out wrong. Thank you for always believing in me.

I would now like to acknowledge my fellow syntax/semantics PhD candidates. Kara Rancourt, you have been a rock throughout this journey. I'm so grateful the stars aligned for us to know each other. Shiyang Fu, you're just simply fantastic. You're the sort of friend who will always make me smile. Lu Jin, you're one of the funniest people I've ever met. I don't think I've ever left a conversation with you without a grin on my face. Finally, Aknan Sindi. You've been unreasonably kind and wonderful. I have thoroughly enjoyed getting to know you. Thank you to you all, my syntax/semantics PhD family. Working with you all has been a blast. I'll cherish our time together in our little office, discussing all things syntax, semantics, and other.

There are many other people at York who have helped, inspired, or comforted me in one way or another, and it would be simply impossible to list them all, and I am aware these acknowledgements are getting ungracefully long. I would however like to formally acknowledge the Department of Language and Linguistic Science as a whole for offering me a fee waiver, without which it would not have been possible to complete this thesis. I would like to also thank the entire admin team in the Department. I really appreciate all the hard work you do, the department would fall apart without you. Also, thank you for putting up with me when I come to distract your very busy days.

There is two linguists from outside of York I need to properly thank. The first is Ryan Walter Smith. You were the first person outside of York who really made me believe in myself and my ideas. You made me feel seen. I really appreciate your insight and discussion. You're also just an all round solid person. Thank you for everything. If I could be half the linguist that you are, I would be incredibly happy. Second, I would like to thank Marcin Wągiel. I actually didn't meet Marcin until the very end of my PhD when I was about to submit my thesis. Until then, he was basically a linguistics-celebrity to me where I had some strange academic-parasocial relationship going on, because I read his thesis so many times. However, when I did finally meet Marcin, he was so unbelievably kind and lovely. Marcin, your confidence and interest in my PhD really spurred me on to finish this thing when I was beyond sick of it. You made me feel like there was at least one person out there who would enjoy what I have to say. Thank you for all the confidence boosts. Thank you for believing in me, it really means a lot.

The final two linguists to thank are my examiners, Louisa Martí and Eytan Zweig. Thank you for taking the time to read this (very long) thesis, and for providing an excellent discussion during the viva. You both made the process unpainful, and your feedback helped me fix the blatantly wrong parts of my proposal. Thank you.

This thesis would also not have been possible without collaboration from my language informants. I am incredibly indebted to everybody who has humoured my insistent questions. Aknan Sindi, Fatma Layas, Latifa Alkuwaiz and Zaki Al-Ghazal, thank you for your time discussing Arabic. Lisa Owen, Siwan Jones, Siôn Pennant Tomos, Gareth Roberts, Megan Hughes, Evan Llewellyn John and Madison K: thank you for your insights on Welsh. Shiyang Fu, Lu Jin, Xiaozhou Zhang, and Jinfei Dong: thank you for your discussions on Mandarin Chinese. Steve Hewitt and Jean Lecoulant: thank you for everything Breton. Purnur Altay and Muhammed Lleri: thank you for discussions of Turkish. And finally, thank you Nina Radkevich for your Russian insights.

Finally, I have a list of friends and family who have kept me sane throughout this journey. First, Ella. Just knowing you're always there for me has been an amazing comfort. I am incredibly lucky to know you, and even luckier to have your unwavering support. Thank you. Olive, your gentle friendship over the years has supported me in ways that are near impossible to describe. Thank you for always picking up the phone even when I'm truly awful at replying to your messages. Jamie and Lauren (or Lauren and Jamie?), you know I loathe to say nice things about you both, so cherish this moment. You have been the greatest source of deep belly laughs, and the best place to vent my frustrations. Thank you both, you're great friends. Katie, one of the best outcomes of this PhD journey was meeting you. Thank you for matching my wannabe goth energy, and feeding my love for all things vampire. Fei, you're truly one of the most unique and fun people I've ever met. Thank you for your support and understanding throughout the years. Francina, I can't describe how much of an inspiration you are. Thank you for always making me want to improve myself. Stew, your energy is unparalleled. Thank you for your humour and keeping me from getting a big head. Siwan, I'm so grateful you have been my cheerleader. Diolch yn fawr iawn. Charlotte, we shouldn't get along. We're too opposite. But here I am, forever thankful to have you and your unwavering support in my life. Finally, Adam, my Olympic lifting coach who has no idea what my thesis is even about. Thank you for keeping me grounded and refusing to let me wallow, even when I broke my ankle.

The last three to thank are my family, only one of which is vaguely human. To Mr. S. Patoumas and Mr. Klaus Cullen, your snuffly noses and adorable paws have been the best stress relief throughout this entire journey. My final acknowledgement is for Oliver, my partner. You've supported me in every step of my educational journey, from my GCSEs all the way up to my PhD. We've grown a lot together. I cannot think of a better person to have been by my side. I am grateful for your love and support, and no amount of thank yous will ever be enough. I'm incredibly excited for the next chapter of my life, and I know you'll be there, supporting me in whatever comes. Thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Where a language exhibits a morphological markedness difference with respect to singular and plural terms, the typical observation is that the singular term is morphologically unmarked, while the plural is morphologically marked (Greenberg, 1972; Corbett, 2000). So called `COLLECTIVE/SINGULATIVE LANGUAGES` have not only the typically expected singular/plural system, but also a minor number system where morphological marking appears on the singular term (a `SINGULATIVE`), and the morphologically unmarked term is number neutral (a `COLLECTIVE`), *i.e.* its meaning refers to both singularities and pluralities. Examples of singulative languages include Welsh (natively, *Cymraeg*) and Arabic (natively, العربية), as shown in Table 1.1.

	Singular → Plural	**Gloss**	**Collective → Singulative**	**Gloss**
Welsh	merch → merch-ed	girl → girls	hwyaid → hwyad-en	duck(s) → duck
Arabic	bint → bin-aat	girl → girls	baṭ → baṭ-ah	duck(s) → duck

Table 1.1: Welsh and Arabic singular/plural and collective/singulative systems

Singulative marking has been known in the typological literature since at least Greenberg (1972),[1] but it wasn't until the 1990s and 2000s that formal scholarship started investigating collective/singulative structures, with notable work from Acuaviva (2008); Corbett (1996); Fassi Fehri (2004); Ojeda (1992) and Zabbal (2002). Collective/singulative systems have gained a surge in interest in recent years, primarily spearheaded by Dali and Mathieu (2021a); Grimm (2012a,b, 2018); Kagan (2024); Kagan and Nurmio (2024); Mathieu (2012a, 2013, 2014); Nurmio (2017, 2023) and Wągiel and Shlikhutka (2023a). Through this boom of research, uncontroversial conclusions regarding the crosslinguistic behaviour of collective/singulative systems have emerged.

First, it is generally accepted that there is some connection between collective/singulative systems and perceptual/construed realities of the world such that terms which fall within the collective/singulative class typically refer to individuals that come in groups, aggregations, swarms, herds, piles *etc.*. As such, the collective/singulative class typically covers the notional categories of grains, fruits, vegetables, insects, and small animals (Asmus and Werner, 2015; Dimmendaal, 2000; Grimm, 2012b, 2018; Haspelmath, 2008; Haspelmath and Karjus, 2017; Hemon, 1975; Roberts and Gathercole, 2012; Haspelmath and Karjus, 2017; Jaradat, 2023a; Jaradat and Jarrah, 2022; Kagan, 2024; Nurmio, 2017; Stolz, 2001; Unseth, 1988; Wągiel and Shlikhutka, 2023a; Wierzbicka, 1988, *a.o.*).

Second, though there is disagreement regarding the underlying mechanisms and theoretical assumptions, collective/singulative systems are (at least descriptively) akin to object mass/general number systems. Specifically, collective terms are akin to

[1]Various languages and language families have been claimed to have a collective/singulative system including: Arbore (Hayward, 1984), Arabic (Dali and Mathieu, 2021a,b; Fassi Fehri, 2004, 2012; Ojeda, 1992; Zabbal, 2002, *a.o.*), Breton (Acuaviva, 2008; Hemon, 2011, 1975; Press, 2004; Stump, 2005), Dagaare (Grimm, 2010), Krongo (Reh, 1985), Majang (Unseth, 1988), Maltese (Acuaviva, 2008; Corbett, 1996; Gil, 1996), Miraña (Seifart, 2005), Murle (Arensen, 1982), Ojibwe Mathieu (2012b), Polish (Asmus and Werner, 2015), Russian (Asmus and Werner, 2015; Kagan, 2024; Wierzbicka, 1988), Sorbian (Asmus and Werner, 2015), Ukrainian (Asmus and Werner, 2015; Kagan, 2024; Wągiel and Shlikhutka, 2022), and Welsh (Grimm, 2012b; Jones and Thomas, 1977; King, 2003; Stolz, 2001; Nurmio, 2017, 2020, 2023, *a.o.*).

object mass and general number in that they denote uncountable multiplicities, while singulative marking is typically taken to be indicative of a function of individuation, classification, or packaging. In this manner, the singulative derives singular count terms from (object) mass, and is classifier-like in function. (Borer and Ouwayda, 2021; Dali and Mathieu, 2021b; Mathieu, 2012a, 2013, 2014; Ojeda, 1992; Grimm, 2012b, 2018; Nurmio, 2017, 2020, 2023; Ouwayda, 2014; Wągiel and Shlikhutka, 2023a; Zabbal, 2002, *a.o.*).

Uncontroversial claims regarding collective/singulative systems:

a. Collective/singulative systems are minor number systems contrasting against a dominant singular/plural system.
b. Collective/singulative systems cover notional classes that have a sense of natural aggregation/clustering.
c. Singulative markers are classifier-like in function.
d. Singulative marked terms are semantically singular.
e. The collective is object mass/general number-like in function.

Though there have been great advancements in the understanding of collective/singulative systems, much of the established collective/singulative work is in silo, with only limited comparison to the singular/plural system of the language (or language family) of enquiry. There is a lack of crosslinguistic comparative work. Despite the widespread understanding that collective/singulative systems are akin to object mass/general number systems, there is very little discussion of collective/singulative systems from the point of view of the count/mass distinction, and there is no detailed and explicit comparison of collective/singulative systems to classifier systems in generalised classifier languages, such as Mandarin Chinese, nor to number marking languages with robust object mass systems, such as English. As far as I am aware, the most detailed crosslinguistic discussion of singulative languages is by Grimm (2012b), but Grimm focusses on the comparison to number marking languages and offers no comparison to generalised classifier languages. The only comparisons of singulative languages to generalised classifier languages that I am aware of are by Nurmio (2023) and Mathieu (2012a, 2013, 2014). Nurmio does not offer a formal analysis, but observes that that the singulative marker may be similar to the the general classifier in Mandarin Chinese. Mathieu does not outright state that collective/singulative systems are object mass/general number systems, however such a conclusion is implicit in their analysis.

In addition, there is a lack of crosslinguistic work which compares singulative languages to each other. At least, there is no work that undertakes such an analysis in any depth. The majority of collective/singulative and collective/singulative adjacent work focuses primarily on one language, or language family. The collective/singulative focus in the literature regards (the dialects of) Arabic (Dali, 2020; Dali and Mathieu, 2020, 2021a,b; Mathieu and Dali, 2021; Fassi Fehri, 2004, 2012; Jaradat, 2023b,a; Mathieu, 2012a, 2013, 2014; Zabbal, 2002; Ojeda, 1992, *inter alia*), though there is some work in Celtic, *e.g.* Welsh (Grimm, 2012b; Nurmio, 2017, 2020, 2023) and Breton (Acuaviva, 2008; Stump, 2005). Recently, there has been emerging scholarship focusing on the (restricted) collective/singulative systems in Slavic languages (Kagan, 2024; Kagan and Nurmio, 2024; Wągiel and Shlikhutka, 2022, 2023a,b,c). There is no detailed work that directly compares singulative languages from diverse language families, at least not at the level of collective/singulative terms in sentential contexts. Much of the discussion of the collective/singulative distinction is at the lexical level, comparing word lists rather than sentential behaviour.

This lack of direct crosslinguistic comparison of singulative languages to each other and to wider number systems is a disservice to the enquiries of nominal number, for three specific reasons which I summarise presently:

- Singulative languages are misrepresented in the literature leading to overgeneralisation.
- Singulative languages offer a unique insight into the locus of the count/mass distinction, a currently overlooked potential connection.
- Singulative languages offer a unique insight into the nature of nominal flexibility, a currently unexplored avenue of enquiry.

1.1 Overgeneralisation

The lack of direct crosslinguistic comparison between different collective/singulative languages has lead to overgeneralisation, particularly regarding discussion of the so-called `PLURAL OF THE SINGULATIVE`. It is well known that the Arabic singulative may pluralise. There are two versions of the Arabic plural of the singulative. The feminine sound plural of the singulative is expressed as feminine suffix -(A)AT is shown by example BAṬ-AH → BAṬ-A-AT (duck.COL-SING → duck.COL-SING-PL = *'duck → ducks'*). The `FEMININE SOUND PLURAL OF THE SINGULATIVE` is the most studied Arabic plural of the singulative form, and is normally associated with paucity in non-numeral contexts (Cowell, 1964; Dali and Mathieu, 2021a,b; Mathieu and Dali, 2021; Jaradat, 2023a,b). The other version of the Arabic plural of the singulative is the `BROKEN PLURAL OF THE SINGULATIVE`, which is less studied, and only acknowledged by Fassi Fehri (2004, 2012). The broken plural of the singulative induces nonconcatenative changes in the stem, and is not associated with paucity in non-numerical contexts. This is exampled by SHAJAR-AH → 'ASHJAAR (tree.COL-SING → tree.COL-SING-BRPL = *'tree → trees'*) Note, not all singulative forms have a broken plural of the singulative form, *e.g.* while SHAJAR-AH has both a feminine sound plural of the singulative and a broken plural of the singulative form, BAṬ-AH only has a feminine sound plural of the singulative.

The existence of the Arabic plural of the singulative is initially surprising: if collective terms such as BAṬ (duck.COL = *'duck(s)'*) and SHAJAR (tree.COL = *'tree(s)'*) are plural-flavoured, why should a morphological process apply to individuate the collective, only to re-pluralise it? In other words, why should BAṬ-A-AT and 'ASHJAAR be attested? There are multiple analyses available for why a singulative may pluralise, including that the plural of the singulative is the 'counting plural' while collectives are (object) mass-like and so cannot combine with numerals (Mathieu, 2012a, 2013, 2014), or that the plural of the singulative is inherently paucal, and so not semantically identical to a collective (Dali and Mathieu, 2020). These conclusions are flawed insomuch as they account only for the Arabic feminine sound plural of the singulative. There is no discussion of the (apparently restricted) broken plural of the singulative, and why it can appear with no special/paucal semantics associated with it.

The other major flaw with current analyses of the plural of the singulative is that there is overgeneralisation of crosslinguistic expectations. The conclusions afforded to the Arabic plural of the singulative are assumed to apply to other singulative languages. There is one glaring problem with this: the Welsh singulative, when individuating pre-theoretical objects, **never** pluralises, a fact that is often misreported (*e.g.* Grimm, 2012b; Dali and Mathieu, 2021b; Stolz, 2001). As the absence of the Welsh plural of the singulative has gone unnoticed, there is (obviously) no current explanation as to why Arabic singulative marked terms may pluralise, but Welsh singulative marked terms must not. The picture of the empirical facts of the plural of the singulative is summarised in table Table 1.2.

	Collective	**Singulative**	**Plural**	**Paucal**	**Gloss**
Welsh	hwyaid	hwyad-en	*	n/a	ducks
	coed	coed-en	*	n/a	trees
Arabic	baṭ	baṭ-ah	-	baṭ-a-at	duck
	shajar	shararah	'ashjaar	shajar-a-at	tree

Table 1.2: Welsh and Arabic plural of the singulative and paucal of the singulative

1.2 The locus of the count/mass distinction

An uncontroversial claim in the literature is that the singulative is classifier-like in nature, and the collective is object mass/general number-like in nature. This is a very high-level generalisation. Looking under the surface, there is a very difficult question to answer: are the primitive systems that underlie collective/singulative systems fundamentally the same as the systems that underlie object mass and/or general number systems? Or are they fundamentally different, but in such a way that similar surface behaviours arise? At a theoretical level, this question concerns the `LEXICALIST/CONSTRUCTIONIST` debate.

In the lexicalist tradition, the approach to nominal number (implicitly) follows Quine (1960)'s wisdom: the count/mass distinction is lexically specified. In this way, there is a pre-syntactic generative component that 'codes' all nouns as either lexically mass or lexically count (Chierchia, 2010b, 1998a,b; Grimm, 2012b; Kagan, 2024; Krifka, 1995; Landman, 2020; Rothstein, 2010, *a.o.*). In the `LEXICO-MEANING` version of lexicalism, the generative component typically incorporates some theory specific interaction between construed ontological distinctions and linguistic representation, such that perceptual realities are the driving force behind count/mass lexical coding. Commonly, the generative component ensures that all liquids, in all languages, will be coded as grammatically mass. Theories differ on the source of object mass terms, with *e.g.* Chierchia (2010b) arguing perceptual vagueness, and Landman (2020) arguing that vertical overlap lead to object mass coding. Whichever theory specific perceptual features are afforded to the pre-syntactic generative component, these features are only applicable in languages with a count/object mass distinction, such as English. For languages with no grammatical count/mass distinction, such as Mandarin Chinese, different mechanisms must be proposed to blanket code all nominal terms as grammatically mass, regardless of perceptual reality. In this manner, there is a fundamental difference between number marking languages and generalised classifier languages at the level of how perceptual realities are integrated into natural language with respect to grammatical coding.

On the other hand, the constructionist approach to nominal number is decidedly anti Quineian: there is no count/mass distinction at the lexical level. All lexical items start life as acategorical root concepts, and all categorical and number features are derived compositionally in the syntax (*e.g.* Bale and Barner, 2009; Borer, 2005a,b; Dali and Mathieu, 2020; de Belder, 2013; Harbour, 2014; Mathieu, 2012a, 2013, 2014; Martí, 2020). In this manner, there is no default nominal coding; count and mass are not nominal categories, but useful descriptive labels for uses of root concepts within certain syntactic frames. Perceptual factors are relevant not for lexical coding, but for regulating appropriate use of syntactic structures in context. Under the constructionist approach, there is no fundamental difference between number marking languages and generalised classifier languages, as the same architecture is responsible for count and mass uses of terms; crosslinguistic differences arise through syntactic means, *e.g.* classifiers in Mandarin Chinese and the plural marker in English are taken to be realisations of the same syntactic projection.

When adopting either a purely lexicalist or purely constructionist approach to nominal number, a decision must be made: should perceptual reality or crosslinguistic parsimony be privileged in the architecture of the system? Singulative languages offer a valuable insight into how this question should be approached, due to the fact that singulative languages occupy theoretical space where it is not immediately clear whether a purely lexicalist or constructionist approach to nominal number is appropriate.

The lexicalist approach for singulative languages

The known connection between construed aggregation and the collective/singulative systems suggests that a meaning based, lexicalist, inspired approach to nominal number may be appropriate. To predict the notional coverage of collective/singulative systems, the lexicalist may want to incorporate construals of aggregation as part of some pre-syntactic generative component, a la Grimm (2012a,b, 2018). However, singulative languages offer a challenge for the nature of crosslinguistic count/mass coding and the identification of semantic primitives.

First, it it is increasingly difficult to explain why perceptual factors affect one language, but not another. This is a general problem for lexicalist approaches to nominal number which is amplified by singulative languages. For example, if both Welsh and Arabic have a generative component influenced by construed aggregation, then why should, for example, AFAL (*'apple'*) be a singular/plural term in Welsh, but TUFAAḤ (*'apple'*) be a collective/singulative term in Arabic? Similarly, why is CACYN (*'wasp'*) collective/singulative in Welsh, but DUBBAR (*'wasp'*) singular/plural in Arabic? How many exceptions are permissible before a generative component based in perceptual reality becomes untenable?

Second, in which ways is the generative component language specific? Clearly for Mandarin Chinese, the generative component cannot include any notion of construed aggregation, as all terms are grammatically mass. But what about English? Is the generative component that determines count/object mass coding fundamentally the same as the generative component in Welsh and Arabic that determines singular/plural or collective/singulative coding?

Even if the perceptual factors related to the generative component can be identified, it turns out that all similarities between collective/singulative and general number systems become entirely circumstantial. In extreme cases, this may lead to an analysis where singulative marking is fundamentally different from classification, forcing an analysis where there are no obvious primitives that underlie their similarities. This is essentially the basic approach and issue with Grimm (2012a,b, 2018): while there is an excellent discussion of the connection of aggregation and nominal number, there is no clear application outside of number marking languages and generalised classifier languages. That is, Grimm's semantics for the singulative marker presupposes aggregation, and therefore cannot be semantically akin to classifiers.

The constructionist approach for singulative languages

Singulative markers are commonly taken to be classifier-like in nature, and collectives are taken to be object mass-like in nature. To capture this parsimony, a constructionist approach may be appropriate. Under a constructionist approach, we may suggest that the singulative marker and classifiers occupy the same functional space in the syntax, and the lack of their projection results in object mass/collective interpretation, a la Mathieu (2012a, 2013, 2014). However, singulative languages offer a challenge for the constructionist regarding the nature of nominal number marking.

It is increasingly difficult to explain exactly which singular count compositions receive singulative marking and those which are for all intents and purposes are zero marked. In other words, if Welsh AFAL (*'apple'*) and HWYAD-EN (*'duck'*) are both categorised by the same syntactic frame, then what accounts for the presence/lack if morphological marking? This is essentially the basic issue for Dali and Mathieu (2021b); Mathieu (2012a, 2013, 2014): while there is crosslinguistic parsimony where classifiers and singulative markers are syntactically and semantically akin, there is no satisfactory restriction on how or when singulative marking should apply.

Singulative languages therefore place us at a crossroads with respect to the count/mass distinction, including a.) the exact manner in which perceptual realities affect linguistic representation, and b.) crosslinguistic parsimony, *i.e.* the reconciliation of similar number marking language, singulative language and generalised classifier language grammatical behaviours. It is prudent, then, to return to the very basics of countability and explore the primitive features that unite object mass, collective and general number systems, and the features which are language specific in an attempt to identify the essential elements a theory of number should offer.

1.3 Nominal flexibility

Any serious investigation into the count/mass distinction includes discussions of nominal flexibility, *i.e.* where 'count' terms can be used in a 'mass' way and vice versa. For example, DOG is typically interpreted in a count way, as in (1). However, in (2) it is interpreted in a mass way, *i.e.* there is a `STUFF-READING` of dog meat in the soup.

(1) There is a dog on the bed English

(2) There is dog in the soup English

Explanations of nominal flexibility are necessarily tied up in both number marking and whether one adheres to a lexicalist or constructionist approach to nominal number. The constructionist approach assumes that all terms are inherently flexible, and the difference between (1) and (2) is due to the root DOG$^{\sqrt{c}}$ appearing in different syntactic frames, and the 'mass' usage of (2) is signalled by the lack of overt number marking (Bale and Barner, 2009; Borer, 2005a; de Belder, 2013). The lexicalist approach on the other hand assumes that nominal flexibility arises through a COUNT → MASS nominal shifting device, which may be triggered by grammatical mismatch of number, *e.g.* when a lexically count noun appears in mass contexts. In this way, DOG in (1) is default, but in (2) there is a covert semantic shift (Cheng et al., 2008; Rothstein, 2010, 2017).

The theories of nominal flexibility are built on the behaviour and interpretation of zero marked 'singular' terms. Singulative languages offer a unique opportunity to explore theories of nominal flexibility: as singulative marked terms are morphologically marked singulars, the constructionist and the lexicalist approaches predict different behaviours for their nominal flexibility. The constructionist approach predicts that singulative marked terms should never be flexible in interpretation, but always be interpreted as singular count, while collective terms may always denote mass 'stuff'. The lexicalist approach should allow for the COUNT → MASS shifting device to apply to singulative marked terms in grammatically incongruent contexts.

To my knowledge, the only investigation into the nominal flexibility of collective/singulative systems is by Kagan (2024), though this work focuses solely on Slavic. While an excellent start for the discussion of nominal flexibility in minor number systems, there is no crosslinguistic comparison outside of the immediate language family. A deeper investigation into the crosslinguistic behaviours of nominal flexibility is required to determine whether the constructionist or lexicalist approach to nominal flexibility is viable.

1.4 Aims and upcoming content

This thesis investigates the collective/singulative distinction across two distinct languages: Welsh, and Arabic. Not only this, but there is explicit comparison to baseline languages with disparate countability properties: English (a number marking language), and Mandarin Chinese (a generalised classifier language). These languages were chosen due to their diverse language genealogy, speaker availability, and the fact that Arabic and Welsh are the most widely spoken singulative languages. Though these are the primary four languages of interest, there are comparisons to other languages, including Breton (another singulative language) and Turkish (another number marking language, with distinct number making properties similar to Welsh). This thesis aims to answer the following questions:

R1 What are the primitive features that underlie collective/singulative, object mass, and general number systems, and what features are unique to each system?

R2 What is the architecture of nominal number?

- a. Is it constructionist in nature, lexicalist in nature, or are these categories too restrictive?
- b. What are the unifying semantics/syntax of object mass, collective and general number terms?
- c. What are the unifying semantics/syntax of singulatives and (general) classifiers?
- d. Why can the Arabic singulative pluralise, but the Welsh singulative cannot?
- e. What can singulative languages tell us about the nature of nominal flexibility?

Though the questions are simple in conception, they require a serious reconsideration of countability in such a manner that questions the nature of the semantics-syntax interface of mass, count, and plurality. To answer these questions, the thesis is organised as follows.

In chapter 2, I explore R1. I compare the collective/singulative systems of Welsh and Hejazi/Modern Standard Arabic to the object mass system of English and the general number system of Mandarin Chinese. I show that these nominal systems share the same primitive semantic features, where object mass, collective and general number are `NUMBER NEUTRAL UNCOUNTABLE` terms. I show that the general classifier in Mandarin Chinese and the singulative marker in Welsh and Hejazi/Modern Standard Arabic are reducible to the same primitive classificatory function. In this chapter I also discuss the status of the plural of the singulative. The plural of the singulative in Hejazi/Modern Standard Arabic behaves as expected: the feminine sound plural of the singulative is paucal while the broken plural of the singulative is a typical plural. The plural of the singulative in Welsh is shown to be absent, a descriptively deviant result.

In chapter 3, I lay the groundwork of how we may think of ontological distinctions such as parts, wholes, substances and objects as pre/extra-linguistic concepts. The distinctions are modelled within a theory of `MEREOTOPOLOGY`, though `DISTANCE FUNCTIONS` are also covered. This chapter does not discuss the linguistic phenomena of this thesis, but the formalisms introduced in this chapter are assumed in later chapters when I discuss/build theories of nominal number.

In chapter 4, I explore R2a, in particular the lexicalist and constructionist approaches to nominal number from the view of collective/singulative languages. The chapter shows that the lexicalist approach (where number features are lexically given) is somewhat successful in capturing something that is intuitively correct: perceptual realities may influence linguistic representation. In particular, Grimm (2012b)'s lexicalist approach is discussed. I suggest that while Grimm's account is successful in capturing that construed aggregation influences nominal number, it nonetheless struggles to account for crosslinguistic facts. This chapter also shows that the constructionist approach (number features are derived in the syntax) has powerful tools to predict crosslinguistic variation. In particular, Mathieu (2012a, 2013, 2014)'s account is successful in that it equates the collective/singulative class to object mass in number marking languages and general number in generalised classifier languages, and the function of the singulative to the function of a classifier. However, in constructionist approaches, there is no clear way to incorporate the manner in which perceptual realities influence linguistic representation. The chapter concludes that a fusion of the strengths of lexicalist and constructionist approaches to nominal number is warranted.

In chapter 5, I address R2b, where I build a theory for nominal number systems. The architecture put forward is in the style of de Vries and Tsoulas (2021)'s `SUBPREDICATIVE ICEBERG SEMANTICS`, which is based on Landman (2020)'s `ICEBERG SEMANTICS`. The Subpredicative Iceberg Semantics account is a hybrid model that affords the strengths of lexicalist and constructionist accounts in that it allows for powerful tools that predict crosslinguistic variation, but in a way such that perceptual realities influence nominal number coding. In this chapter, I offer an adaptation to Subpredicative Iceberg Semantics that allows for singulative languages to be incorporated into the model such that construals of aggregation trigger number neutral uncountable nominal derivation in collective/singulative languages.

In chapter 6, I expand the Subpredicative Iceberg Semantics account as to address R2c and R2d. A formal semantics is proposed for the individuating singulative (which is taken as an uncountable classifier), the packaging singulative, countable classifiers, countable plurals, and uncountable plurals. The semantics is worked in a way such that it is compatible with a split analysis of nominal number, *i.e.* number features are not restricted to a single syntactic projection. The plural of the singulative is taken to be a structure that is derived via stacking functions of number across the nominal spine. The semantics are worked out such that the plural of the singulative is predicted in Arabic, but ruled out in Welsh on grounds of vacuousness.

In chapter 7, I explore incorporating nominal flexibility into the Subpredicative Iceberg Semantics model in line with R2e. Through exploration of Welsh, Hejazi

Arabic, English and Mandarin Chinese, I show that fruit and vegetable terms across languages allow stuff-readings if the context is amenable. Further, singulative marked terms in Welsh (but not Hejazi Arabic) allow stuff-readings if the context is amenable, a fact not predicted by either a pure last resort approach or root meaning approach approach to nominal flexibility. I offer a multiple-source analysis of nominal flexibility. One source of nominal flexibility is `PERSPECTIVE SHIFT`, which is taken to be universal. Perspective shift concerns the interaction of grammar and construed judgements of substance and object. The second source of nominal flexibility is a COUNT → MASS shifting device, GRIND, which applies to singular count and plural terms. The application of GRIND differs crosslinguistically, and as such it is neither universal not a last resort mechanism (*c.f.* Cheng et al., 2008; Rothstein, 2010, 2011). The third source of 'nominal flexibility' is due to what I call `ACCIDENTALLY HOMOPHONOUS ROOT CONCEPTS`, which is taken to be a case where two conceptually related properties have distinct grammatical representations but identical phonological realisations, rising to the illusion of nominal flexibility.

Chapter 8 summarises the thesis, tying together to conclusions for R1 and R2a-e. Further avenues for future research are highlighted.

1.5 Conventions

The conventions used in this thesis are more or less standard, though I present the following for clarity.

1.5.1 Dialectal variation

The forthcoming discussion will focus primarily on the dialects of Hejazi Arabic (spoken in Saudi Arabia), and Modern Standard Arabic (as it is spoken in Saudi Arabia). Reference to other dialects of Arabic will be made. The text is explicit regarding which dialect of Arabic will is refereed to in any given discussion. The use of the term `ARABIC` where no dialect is mentioned is used in two cases. The first case is where I am paraphrasing other authors and it is not clear which dialect the source material is referring to. The second case is where the features discussed are assumed to apply to most, if not all, dialects of Arabic.

In Welsh, there is dialectal variation between speakers of Northern Welsh and Southern Welsh dialects, primarily in word and spelling choices. The data in this thesis primarily follows the standard Welsh variety spoken in North Wales. Unless stated otherwise, examples in this these are in Northern Welsh.

1.5.2 Orthography

This work presents original data from Welsh, Arabic and Mandarin Chinese. The presentation of these languages in written format requires contextualising so that examples are not misunderstood. All original examples were created in consultation with native speakers, but any spelling/transliteration errors are entirely my own. Any examples that come from literature are, unless stated otherwise, presented verbatim as they appear in their original source (and thus may be inconsistent with the orthographical standards outlined here).

Welsh is a language with PHONOLOGICAL MUTATION. This is where the phonology/shape of the initial consonant of a word changes in certain grammatical contexts, including but most definitely not limited to: after a preposition, when the object directly follows the subject in canonical VSO order, and after complement marker YN (see King, 2003, for an overview on Welsh mutation environments). Phonological mutation has no semantic effect in and of itself, but is a reflection of grammatical function. The original sound before mutation has occurred is known as the RADICAL form. Table 1.3 shows examples of radical forms and their SOFT MUTATION forms, the most common mutation seen in this thesis. Nasal and aspirate mutation are not shown in Table 1.3 as their relevant appearances in this thesis are very limited.

Phonological mutation is represented in Welsh orthography, and as such is present in the examples in this thesis where appropriate. I do not signal in the examples where mutation has taken place, as this is not the focus of this thesis. Note, in-text examples are isolated, and so will always appear in their radical form, even if the corresponding in-context example exhibits mutation.

Soft mutation	**Radical Form**	**Soft Mutation Form**	**Gloss**
p → b	pump	bum(p)	five
m → f	merch	ferch	girl
g → ∅	gellyg	ellyg	pear(s)
c → g	caws	gaws	cheese
d → dd	darn	ddarn	piece

Table 1.3: Welsh soft mutation

Another important aspect of Welsh orthography is the form of the definite article. Following standard written Welsh conventions, the definite article is written as Y when followed by a consonant (this also triggers soft mutation for following singular feminine nouns). When used before a vowel, the definite article is written YR. When following a vowel, the definite article is written 'R and is attached to the previous word, regardless of whether the following word begins or ends with a vowel.

Both Arabic and Mandarin Chinese are not typically written with the Roman script as Welsh or English is. As such, explicit choices have been made regarding the presentation of these languages. The original Mandarin Chinese examples in this thesis are presented using (an adapted version of) Pinyin.

The original Arabic examples are presented using a semi-non-standard transliteration method. The method I use is based on that used by Haywood and Nahmad (1965, p.3-4), with minor alterations for the long vowel sounds (ا و ي) which I have chosen to represent as letter doubling for reasons relating to the transcription of feminine sound plurals (which I discuss in chapter 2). Note, I only show the isolated form of consonants in Table 1.4, and for reasons of brevity I do not show my transliteration methods for harakaat (short vowel sounds). I have added the formal feminine ending, taa mabuuta ة, to the table, as this is the suffix used to express feminine endings, including singulatives, when no case marking is present. Two versions of taa mabuuta transliteration are presented: one where it is not followed by further morphological material (ah), and a version where it is followed by further morphological material, such as case endings, (a). Note, for consistency, the transliterated of Arabic examples in this work are with respect to the the written version of sentences, rather than reflecting an exact pronunciation of the Hejazi Arabic dialect.

ا	aa	س	s	ل	l
ب	b	ش	sh	م	m
ت	t	ص	ṣ	ن	n
ث	th	ض	ḍ	ه	h
ج	j	ط	ṭ	و	w,uu
ح	ḥ	ظ	ẓ	ى	y, ii
خ	kh	ع	3	ء	’
د	d	غ	gh	ة	ah / a
ذ	dh	ف	f		
ر	r	ق	q		
ز	z	ك	k		

Table 1.4: Arabic transcription system

1.5.3 Examples and glossing

This work is comparative in nature. To avoid confusion, all in-text examples are presented in SMALL CAPS. As the nature of this thesis deals with singulative suffixes, dashes are used to separate the suffix from the base, even if it is not typically represented that way in the target language's orthographical system. If the meaning is incorporated into the base, a dot is used. The following is an example of an in-text example: English CARROT and Welsh MORON-EN (carrot.COL-SING = *'carrot'*).

In-line examples are always labelled with their language of original, even if it is English. To mark outright ungrammaticality, I opt to use asterisk notation, *, as in (3). To indicate awkwardness I use hash notation, # as in (4).

(3) *This is ungrammatical an sentence — English

(4) #Three bloods please — English

The glossing conventions that I will use throughout this thesis are listed in Table 1.5.

1	first person	M	masculine
2	second person	NEG	negation
3	third person	NOM	nominative
ACC	accusative	NNU	number neutral uncountable
CL	classifier	PAU	paucal
CL_{SING}	Classifier (singular)	PL	plural
CL_{PLUR}	Classifier (plural)	PL_{NC}	plural (non-countable)
COL	collective	PL_{C}	plural (countable)
COP	copula	PSING	packaging singulative
DEF	definite	PST	past tense
DEM	demonstrative	SG	singular
F	feminine	TOP	topic
GEN	genitive	SMASS	stuff mass
IMP	imperative	SING	individuating singulative
INDEF	indefinite	YN	Welsh complement marker
JUSS	Jussive mood		

Table 1.5: Glossing system

Chapter 2

Crosslinguistic Comparisons

This chapter offers a descriptive overview of the count/mass and collective/singulative systems of Welsh and Hejazi/Modern Standard Arabic, with explicit comparison to the singular/plural and count/mass systems of English and the general number system of Mandarin Chinese.

At their most basic level, the Welsh and Hejazi/Modern Standard Arabic collective/singulative systems are taken to be minor number systems with a morphologically marked singular, a reversal of the typical expectation of number marking. In the forthcoming discussion, I show that there is nothing special about collective/singulative systems from the point of view of the count/mass distinction, other than the fact that the classifier is realised as a suffix (*i.e.* the singulative marker) rather than a stand alone morph and that singulative languages are reducible to number marking languages where the collective/singulative system is a highly productive object mass system. While there is nothing inherently special regarding the grammatical function of the singulative from the view of the count/mass distinction, the behaviour of the so-called plural of the singulative is a puzzle. I show that the Hejazi/Modern Standard Arabic plural of the singulative facts are expected, but the Welsh plural of the singulative is completely absent. Explicitly, the fact that there is no Welsh plural of the singulative is not predicted.

To this end, in §2.1, I set the scene of enquiry. I overview the number marking facts for the singular/plural and collective/singulative systems of Welsh and Arabic. In this section, it emerges that the term `SINGULATIVE` should be disambiguated into its different uses, which includes at the very least a classifier (`INDIVIDUATING SINGULATIVE`) and a portioning device (`PACKAGING SINGULATIVE`). Other types of singulative are introduced, such as the `FAUX SINGULATIVE`, which is a suppletive singular/plural pair which looks as if it were a collective/singulative pair, but it is not.

In §2.2, I investigate the collective/singulative system from the point of view of the count/mass distinction. This section contains explicit comparison of the Welsh and Hejazi/Modern Standard Arabic collective/singulative systems to the object mass system in English and general number system in Mandarin Chinese. It is shown that collective/singulative systems, object mass systems and general number systems are reducible to the same primitive features: `NUMBER NEUTRALITY`, and `UNCOUNTABILITY`. Further, the individuating singulative marker and general classifiers are shown to be reducible to the same basic classificatory function.

In §2.3, the status of the plural of the singulative is investigated. It is often claimed that the plural of the singulative in Arabic has a paucal interpretation only, while the Welsh plural of the singulative is a regular plural. I show that both these claims are contrary to fact. The Hejazi/Modern Standard Arabic feminine sound plural of the singulative is shown to be paucal while the broken plural of the singulative is a typical plural. The (overt) Welsh plural of the singulative is shown to be completely absent, contrary to reasonable prediction.

Before starting this chapter, the reader is encouraged to review §1.5.2 on Welsh orthography (specifically on phonological mutation), so that no confusion is made when in-text and numbered examples are not orthographically identical.

2.1 Setting the scene

2.1.1 Number marking in Welsh and Arabic

Collective/singulative systems are minor number systems which contrast with a dominant singular/plural system. There is no known language which has a collective/singulative system as its only system (Corbett, 1996).[1] Therefore, a discussion of collective/singulative systems is incomplete without an understanding of the singular/plural systems to which they contrast.

Consider first Welsh. As shown by Jones and Thomas (1977) and King (2003), Welsh has multiple plural forming strategies, including a myriad of suffixes, internal vowel change, a mix of both suffixation and vowel change, and suppletion. There is no semantic difference with respect to the phonological realisation of the plural marker. Examples of each plural marking method are shown in Table 2.1.

	Singular	**Plural**	**Gloss**
	dyn	dyn-ion	man → men
	afal	afal-au	apple → apples
Suffixation	merch	merch-ed	girl → girls
	beic	beic-iau	bike → bikes
	llew	llew-od	lion → lions
Vowel Change	bard	beirdd	poet → poets
Suffixation + Vowel Change	drws	drys-au	door → doors
Suppletion	mats-en	mats-ys	match → matches

Table 2.1: Welsh plural marking strategies

[1] As I ultimately take collective/singulative systems to be classifier systems, I take the stance that there are languages with a collective/singulative only system, but we just call them by another name: generalised classifier languages.

Moving onto the Arabic singular/plural system. It is well known that the varieties of Arabic have varied methods of indicating plurality, including: `SOUND PLURALS`, where a gendered suffix is used; `BROKEN PLURALS`, a nonconcatenative system where there is internal modification of the noun; and gendered `DUAL` suffixes where plurality is limited to a numerosity of *two*. Of these methods, two are of interest to us: the sound plural, and the broken plural.

As discussed by McCarthy and Prince (1990), the predominant method of plural marking in Arabic is the broken plural. In the broken plural system, singular stems are mapped onto syllabic templates to form the plural. Wright (1896, §304) lists thirty-one broken plural templates of varying productivity in Classical Arabic, which McCarthy and Prince categorise into four shape-defined categories. For example, singular forms with a consonant-vowel-consonant-consonant (CvCC) structure are typically mapped to a consonant-vowel-consonant-long.vowel-consonant (CvCvvC) structure when pluralised. The root consonants do not typically change in broken plural formation, but the vowel may do so. Examples of broken plural formation patterns are shown in Table 2.2.

Singular Pattern	**Singular**	**Plural**	**Gloss**
CvCC	nafs	nufuus	soul → souls
	qidh	qidaah	arrow → arrows
CvCvC	’asad	’usuud	lion → lions
	rajul	rijaal	man → men
CvCvvC + at	sahaabat	sahaa’ib	cloud → clouds
	jazzerat	jazaa’ir	island → islands

Examples from McCarthy and Prince (1990, p.217)

Table 2.2: Arabic broken plural patterns

The second Arabic pluralising method of interest is the `FEMININE SOUND PLURAL`, which is realised as a suffix -AAT (ات). In some cases, the feminine sound plural causes minor changes in the stem. The feminine sound plural can in theory apply to any singular nominal in Arabic. However, it typically attaches to (with some exceptions): borrowed terms; nominals that denote female social gender (be those where feminine gender is derived via suffixation or inherent to the noun); singular nominals that end in the formal feminine termination -AH (ة), when the singular form has no broken plural form, in paucal environments, and as a plural of all diminutives.[2] Examples of the feminine sound plural are shown in Table 2.3, where it should be noted that terms ending in -AH, whether derived or inherent, realise the sound plural by fusing the feminine ending with vowel lengthening.[3] To make this clear in forthcoming glosses, in cases where the feminine ending is derived before being pluralised, I will represent the plural as -A-AT as in ZAWJ-A-AT (spouse-F-PL = *'wives'*). This is so the plural is still represented as a long vowel sound, but it is clear that the singular was a derived feminine term. In the case that the feminine feature was due to a noun naturally ending in AH, I will simply represent the plural as -AT with no indication of the absorption of a derived feminine suffix, but absorption of natural feminine endings as in DARAJA-AT (bike.SG-PL = *'bikes'*).

	Singular	**Plural**	**Gloss**
Borrowed term	saandwiitsh	saandwiitsh-aat	sandwich → sandwiches
	tilifizyuun	tilifizyuun-aat	television → televisions
Human female (Inherent)	fatah	fatay-aat	girl → girls
	bint	ban-aat	girl → girls
Human female (Derived)	ṭalib-ah	ṭalib-a-at	student → students
	zawj-ah	zawj-a-at	wife → wives
Nominals ending in -ah	darajah	daraja-at	bike → bikes
	taawilah	taawila-at	table → tables

Table 2.3: Arabic feminine sound plural marking

[2]See Wright (1896, §303, p. 197-99) for a more complete list of where the feminine sound plural may apply, and Alshboul et al. (2013); Lahrouchi and Ridouane (2016) for information regarding the sound plural as a 'default' plural for diminutives.

[3]In older grammars, such as Haywood and Nahmad (1965), the feminine ending is often described as being removed before the addition of the sound feminine plural.

Though morphological marking of the plural term is the typical expectation, collective/singulative languages observe a reverse marked minor number system in the sense of Corbett (2000).[4] Explicitly, this is the case where morphological marking appears on the singular term, while the unmarked term is plural in nature. Throughout the literature, the morphologically marked singular is regularly called a `SINGULATIVE`, and the term from which it is derived is often called a `COLLECTIVE`.

For Arabic, singulative marking takes the form of the feminine singular -AH which induces gender shift on the head noun. Note, singulative marking has some phonological variants, including -AT when pronounced before case endings. In Welsh, singulative marking agrees in gender with the head noun -YN (masculine) and -EN (feminine). Select examples of the collective/singulative class are given in Table 2.4 for Welsh and Table 2.5 for Arabic, though the reader is directed to Appendix A for a more comprehensive list, where over 100 pairs of collective/singulative terms have been identified in both Welsh and in Arabic.

Collective	**Singulative**	**Gloss**
pysgod	pysgod-yn	fish(es) → a fish
moch	moch-yn	pig(s) → pig
malwod	malwod-yn	snail(s) → snail
moron	moron-yn	carrot(s) → carrot
gellyg	gellyg-en	pear(s) → pear
mogrug	mogrug-yn	ant(s) → ant
ceich	ceirch-yn	oat(s) → oat
cnau	cnau-en	nut(s) → nut
afen	afen-en	raspber(ies) → raspberry
coed	coed-en	tree(s) → tree
ysgaw	ysgaw-en	elder tree(s) → elder tree
hwyaid	hwyad-en	duck(s) → duck

Table 2.4: Welsh collective/singulative marking

[4] As a minor number system, collective/singulative systems limited coverage. For example, a corpus measure by Roberts and Gathercole (2012) indicates that collective/singulative systems account for 2.5% of nouns in Welsh.

Collective	**Singulative**	**Gloss**
samak	samak-ah	fish(es) → a fish
naml	naml-ah	ant(s) → ant
jazar	jazar-ah	carrot(s) → carrot
baT	baT-ah	duck(s) → duck
shajar	shajar-ah	tree(s) → tree
naxal	naxl-ah	palm tree(s) → palm tree
tufāh	tufāh-ah	apple(s) → apple
beyd	beyd-ah	egg(s) → egg
tamr	tamr-ah	date(s) → date
burtagal	burtagal-ah	orange(s) → orange
futr	futr-ah	mushroom(s) → mushroom
baqar	baqar-ah	cow(s) → cow

Table 2.5: Arabic collective/singulative marking

A note on terminology is appropriate here. I will follow tradition and opt to use the term `COLLECTIVE` due to its historical significance and regular use in the literature (Such as Acquaviva, 2015; Dali and Mathieu, 2020, 2021a,b; Mathieu and Dali, 2021; Fassi Fehri, 2012, 2016, 2020; Greenberg, 1972; Grimm, 2012a,b; Mathieu, 2012a,b, 2013, 2014, 2018; Nurmio, 2017, 2023; Wągiel and Shlikhutka, 2023a, *a.o.*).[5] For the purposes of this thesis, the term `COLLECTIVE` in this sense is a convenient label, and as Nurmio (2017, p. 60) before me also stresses, the use of this term does not offer any theoretical significance and is not to be confused `LEXICAL COLLECTIVES`[6] such as GROUP, COMMITTEE *etc.* in a Landman (1989a,b) sense, nor `MORPHOLOGICAL COLLECTIVE` in the Grimm and Dočekal (2021) sense, where collectives are derived via suffixation. Secondly, I will use the term `COLLECTIVE/SINGULATIVE SYSTEM(S)` interchangeably with `COLLECTIVE/SINGULATIVE CLASS`. The choice of which term used is not significant, and is done only to aid readability.

[5]Though *c.f.* Borer and Ouwayda (2021) who refer to collective terms as `BATCH NOUNS`.
[6]Also known as a `SOCIAL COLLECTIVE` in Wągiel and Shlikhutka (2023a).

2.1.2 Individuating and packaging singulatives

The discussion so far may give the illusion that the term `SINGULATIVE` is only used to describe a marked singular term contrasting against a collective, which is understood as an unmarked plural-flavoured term. As it turns out, there is no consensus on the term `SINGULATIVE` as multiple distinct semantic operations have in recent years fallen under its remit. This state of affairs was pointed out by Acquaviva (2015), who stated that the term `SINGULATIVE` has both broad and narrow use. In its narrow sense, it is used to define a unitisation operator. In its broad sense, it is used not only to define a unitisation operator, but also a packaging/portioning operator. The broad sense is the generally adopted definition. For example, Nurmio (2023) allows singulatives to be defined not only as a unitisation operator upon a collective, but also as a packaging operator, *i.e.* creating individuals from a substance mass term:

> A singulative is a derived noun form which is formed by adding a marker to a non-unit-denoting base and which denotes 'a/one X' or 'a/one unit of X'
> (Nurmio, 2023, p.2)

Similar observations are seen in the Arabic literature. For example, Ojeda (1992) states that singulatives:

> ...are derived from collectives and refer either to 'a specific quantity of the substance', or else 'an individual member of the collection'.
> (Ojeda, 1992, p.307)

That broad sense of singulative marking seems apt to describe the Welsh system, where the singulative suffix -YN/-EN has two different interpretations, dependent upon whether the base it attaches to denotes objects or stuff.[7] The Welsh collective term GELLYG (pear.COL = *'pears'*) in its normal use denotes a singular/plural agnostic

[7] A similar distinction is pointed out by Grimm (2012b, 2018) for Maltese, in a footnote of Grimm (2018, p.15) for Dagaare, and by Mathieu (2012a) for Ojibwe and Breton.

numerosity of pears, but each pear is an integrated whole.[8] When applied to GELLYG, the singulative marker in (1) derives a singular term denoting an integrated whole, in this case a whole, singular pear. Importantly, the denotation of the collective and singulative both have reference to the same types of wholes; the difference between the collective and the singulative is simply of numerosity. In a sense, the singulative can be said to be 'picking out' an individual member from a collection or a group of like-individuals. This is the function of what I call the `INDIVIDUATING SINGULATIVE`, which I gloss as SING.

(1) Bwytodd Owain **ellyg-en** ddoe — Welsh
eat.PST.3 Owain pear-SING yesterday
'*Owain ate a pear yesterday*'

(2) `INDIVIDUATING SINGULATIVE`
A suffix which derives a singular from a collective. It 'picks out' a predefined singular from a plurality.

Consider now the term CAWS (cheese.SMASS = '*cheese*'), which is a canonical stuff mass term. When the singulative marker applies to CAWS to create COS-YN there is a part-of/packaging/portioning effect in the Bunt (1985); Chierchia (1998a, 2010b) manner, as seen in (3) where the predominantly mass reading is made count. In this case, singulative marking does not 'pick out' pre-theoretical individuals, but rather *creates* an individual from the mass, a portion/piece of cheese[9]. Therefore, in this case, the difference between the singulative and non-singulative term is not of numerosity, but closer to a material part-whole distinction. This is the function of what I call the `PACKAGING SINGULATIVE`, which I glass as PSING.

(3) Bwytodd Owain **gos-yn** ddoe — Welsh
eat.PST.3 Owain cheese-PSING yesterday
'*Owain ate a piece of cheese yesterday*'

[8] In some contexts, GELLYG may mean something close to PEAR STUFF as a mass. This will be discussed at length in chapter 7.

[9] *c.f.* Stolz (2001) who claims that the singulative in COS-YN is a diminutive marker.

(4) PACKAGING SINGULATIVE
A suffix with a packaging function. It creates portions from mass terms.

Further support that singulative marking has two distinct functions/interpretations is that only packaging singulatives are interchangeable with DARN (*'piece'*) if a part-whole meaning is to be maintained. Explicitly, DARN is semantically equivalent only to the packaging singulative meaning: the difference between (6) and (3) is purely stylistic. On the other hand, the difference between (5) and (1) is in the type of function used: the interpretation of (5) is portioning, while the interpretation of (1) is individuating.

(5) Bwytodd Owain **ddarn** o ellyg ddoe — Welsh
eat.PST.3 Owain piece of pear yesterday
'*Owain ate a piece of pear yesterday*' $\neq$ (1)

(6) Bwytodd Owain **ddarn** o gaws ddoe — Welsh
eat.PST.3 Owain piece of cheese yesterday
'*Owain ate a piece of cheese yesterday*' $=$ (3)

The productivity of individuating singulatives and packaging singulatives are not equal in and across languages. While the individuating singulative may apply to almost any collective, the packaging singulative is not always productive. Nurmio (2023) points out in a footnote that singulatives on mass terms in Welsh are highly lexicalised. Nurmio argues that in the case of COS-YN the use of the packaging singulative may only refer to pre-packaged cheese (such as the 250g block of cheddar in your local supermarket), rather than simply 'A PIECE OF CHEESE'. While it is true that some packaging singulatives may be highly lexicalised, this is not the case for all packaging singulatives. The examples in Table 2.6, in context in (7), do not refer to pre-packaged individuals, but may only refer to portions in context.[10]

[10] While CACEN-EN (cake.SMASS-PSING = *'a piece of cake'*) is possible, it is dispreferred, due to its phonological similarity to singulative term CACYN-EN (wasp.COL-SING = *'wasp'*). Rather, the construction using DARN (*'piece'*) is preferred, *e.g. Bwytodd Owain darn o gagen ddoe.*

Mass	**Packaging singulative**	**Gloss**
menyn	menyn-en	butter → a piece of butter
cig	cig-yn	meat → a piece of meat
cacen	cacen-en	cake → a piece of cake

Table 2.6: Welsh packaging singulatives

(7) Bwytodd Owain **gig-yn** ddoe — Welsh
eat.PST.3 Owain meat-PSING yesterday
'*Owain ate a piece of meat yesterday*'

In Arabic, the productivity differences between the individuating and packaging singulatives are obscure. Clearly, the varieties of Arabic have an individuating singulative: every singulative in Table 2.5 is an individuating singulative, and (8) and (9) exemplify for Hejazi Arabic and Modern Standard Arabic, respectively.[11]

(8) Malik 'akala tufaaḥ-**ah** 'ams — H. Arabic
Malik eat.PST.M.3 apple-SING yesterday
'*Malik ate an apple yesterday*'

(9) 'akala Malik-u tufaaḥ-**at**-an bil'ams — MSA
eat.PST.M.3 Malik-NOM apple-SING-ACC yesterday
'*Malik ate an apple yesterday*'

The status of the Arabic packaging singulative is not entirely clear. For Greenberg (1972) and Ojeda (1992) there is a packaging singulative in (Classical) Arabic, citing examples such as XISHA-AH (wood.SMASS-PSING = *'piece of wood'*), JUBN-AH (cheese.SMASS-PSING = *'piece of cheese'*) and ZUBD-AH (butter.SMASS-PSING = *'portion of butter'*) as evidence. In modern Arabic vernaculars, the packaging singulative is

[11] Canonically Arabic is a VSO language. However, in dialect, including in Hejazi Arabic, alternation between VSO and SVO is common. The examples for positive existential sentences in Hejazi Arabic in this thesis are the SVO forms. The examples in Modern Standard Arabic maintain canonical VSO structure.

variably argued for. Jaradat and Jarrah (2022) argue that in Jordanian Arabic JUBN-AH is not an instance of a packaging singulative applying to a canonical mass term. Their argument is that JUBN-AH may appear with other dividers, *e.g.* QITE-A JUBN-AH (piece-F cheese-F = *'piece of cheese'*), and therefore -AH is better analysed a feminine agreement marker.

Jaradat and Jarrah's feminine agreement marker analysis may not be appropriate for all modern vernaculars, where the packaging singulative is reported. Borer and Ouwayda (2021) report that mass term BIIRAH (*'beer'*) can be unitised as BIIRAY-EH (beer.SMASS-PSING = *'portion of beer'*) in Lebanese Arabic. Similarly, Hnout et al. (2021) report a packaging singulative in the Galilee dialects of Palestinian Arabic, *e.g.* XOBE (*'bread'*) is unitised as XOBZ-E (bread.SMASS-PSING = *'piece of bread'*). For the dialect of interest in this thesis, Hejazi Arabic, while there is a preference to use portion terms, such as QITEA (*'piece'*), the packaging singulative may be marginally used, but speakers report the sentence as feeling jocular, or something that may be said to children. It would then seem that the use of the packaging singulative in Hejazi Arabic is more free than in Jordanian Arabic, but it is still not as productive as was claimed by Ojeda for classical Arabic. In MSA the pattern is more similar to Jaradat and Jarrah's report of Jordanian Arabic, *i.e.* speakers assert a preference to use a free morph rather than the singulative for packaging interpretations. From this data, we may tentatively conclude that there is a packaging singulative present in some (but perhaps not all) dialects of Arabic, but it is not entirely productive.[12]

(10) Malik 'akala jubn-**ah** 'ams — H. Arabic
Malik eat.PST.M.3 cheese-PSING yesterday
Marginally: '*Malik ate (a piece of) cheese yesterday*'

(11) #'akala Malik-u jubn-**at**-an bil'ams — MSA
eat.PST.M.3 Malik-NOM cheese-PSING-ACC yesterday
Intended: '*Malik ate a piece of cheese yesterday*'

[12] Hnout et al. (2021) report that the packaging singulative in Galilee Palestinian Arabic is not fully productive for all types of stuff mass terms. According to Hnout et al., the Palestinian packaging singulative may only apply to solid stuff mass terms, and not liquid stuff mass terms.

Asymmetry in the productivity of individuating and packaging singulatives may well be the normal state of affairs. This observation has been seen in various unrelated works on the collective/singulative systems of Slavic languages. Wierzbicka (1988, p 518) claims Russian has only an individuating singulative, citing the examples such as in Table 2.7.[13] Kagan and Nurmio (2024) also suggest there is nuance to the Russian packaging singulative, suggesting not all packaged meanings are alike. In Ukrainian, Wągiel and Shlikhutka (2022, 2023a) argue that the packaging singulative may only apply to substance mass terms if they denote liquids that are portioned into relevant instances, and that in general solid substance mass terms do not combine with the packaging singulative. For example, the terms in Table 2.7 are said to be valid instances of the Ukrainian packaging singulative.

Language	**Collective/Mass**	**Singulative**	**Gloss**
Russian	gorox	goroš-ina	pea(s) → pea
	fasol	fasol-ina	bean(s) → bean
	mjaso	*mjas-ina	meat → *piece of meat
Ukranian	rosa	ros-yna	dew → dew drop
	došč	došč-yna	rain → rain drop

Russian examples from Wierzbicka (1988)
Ukrainian examples from Wągiel and Shlikhutka (2022)

Table 2.7: Slavic collective/singulative marking

Considering the Slavic evidence alongside the Welsh and Hejazi/Modern Standard Arabic evidence, it would seem that both packaging singulative and individuating singulative distributions are treated differently in and amongst languages. The most common pattern, as seen in Arabic dialects and Slavic languages, is that the packaging singulative is more restricted than the individuating singulative.

[13] Wierzbicka (1988, 518) claims that RIS → *RIS-INA (rice.COL → rice.COL-SING = *'rice → *grain of rice'*) is not attested in Russian, but according to Nina Radkevicj *p.c.* RIS → RIS-INA (rice.COL → rice.COL-SING = *'rice → grain of rice'*) is a valid collective/singulative pair. I put this down to speaker variation. As we will see in §4.2 speaker variation is a general feature of collective/singulative systems.

In this section I have covered the broad and narrow use of the term `SINGULATIVE`. The broad use of the term `SINGULATIVE` to cover both individuating and packaging singulative operations is often intentional. For Ojeda (1992), the conflation of the Arabic individuating and packaging singulatives allows a unified semantics to account for both interpretations under a refinement function.[14] Though different in execution, a similar approach is taken Mathieu (2012a, 2013, 2014) and Dali and Mathieu (2021a). In these accounts, individuating and packaging singulatives receive the same formal treatment, *i.e.* they are both derived under the same syntactic head in a Borer (2005a) inspired fashion. In contrast, Grimm (2012b) uses the term `SINGULATIVE` to refer to both individuating singulatives and packaging singulatives, but assumes different formal treatments for each function. Specifically, Grimm proposes a function that presupposes defined individuals to analyse the Welsh individuating singulative, essentially excluding packaging singulatives from its semantics.[15] A contention for this thesis is that the term `SINGULATIVE` should be disambiguated, at least at the descriptive level. The conflation of senses for the term `SINGULATIVE` has lead to inaccurate claims regarding the Welsh collective/singulative system throughout the literature that are taken as fact, nominally the status of the plural of the singulative, a matter to which I return to in §2.3.2.

2.1.3 Artificial collection singulatives

In Welsh, there is a candidate for another type of singulative, which I call the `ARTIFICIAL COLLECTION SINGULATIVE`, which attaches to collectives that denote artificial collections such as DODREFN (*'furniture'*). The artificial collection singulative is clearly comparable to the individuating singulative as 'picks out' a predefined individual (*i.e.* and object) from a collection, but it differs in a specific way: only artificial collective singulatives are interchangeable with DARN (*'piece'*) where there is no tangible difference in meaning. For example, (13) and (14) have

[14]Kagan (2024) also provides a unifying semantics for the Slavic individuating singulative and packaging singulative which is based on partitioning.

[15]While recognising that the packaging singulative exists, Grimm (2012a) offers no formal treatment for it.

the same meaning, and as already mentioned, (1) and (5), repeated below, differ. In this thesis, I will have very little to say about artificial collection collectives, and I present the data here only for a more holistic picture.

(12) Prynodd Owain dodrefn ddoe — Welsh
buy.PST.3 Owain furniture yesterday
'*Owain bought furniture yesterday*'

(13) Prynodd Owain dodrefn-**yn** ddoe — Welsh
buy.PST.3 Owain furniture-SING yesterday
'*Owain bought a piece of furniture yesterday*' = (14)

(14) Prynodd Owain **ddarn** o dodrefn ddoe — Welsh
buy.PST.3 Owain piece of furniture yesterday
'*Owain bought a piece of furniture yesterday*' = (13)

(1) Bwytodd Owain ellyg-**en** ddoe — Welsh
eat.PST.3 Owain pear-SING yesterday
'*Owain ate a pear yesterday*'

(5) Bwytodd Owain **ddarn** o ellyg ddoe — Welsh
eat.PST.3 Owain piece of pear yesterday
'*Owain ate a piece of pear yesterday*' ≠(1)

2.1.4 Non-physical singulatives

So far, all collective/singulative examples given have denoted physical entities, either conceived as natural kind concepts or artificial collection concepts. There is a reasonable question regarding whether the collective/singulative class covers non-physical and/or abstract concepts. To the best of my knowledge, the collective/singulative class in Welsh covers only physical entities, as discussed above. In Arabic, the picture is not so clear. Fassi Fehri (2012), for example, reports that an activity such as RAQASA (*'dance'*) can denote a plurality of events when used verbally, as in (15), but when used nominally and paired with the suffix -AH, it denotes a singular event, as in (16), where I borrow Fassi Fehri's glossing convention of the 'singulative' as 'unit'.

(15) **raqasa** r-rijaal-u thalaath-a marraat-in — Arabic
danced DEF-man.PL three-ACC times.PL-GEN
'*The men danced three times*'
From (Fassi Fehri, 2012, p.309)

(16) raqasa r-rijaal-u 'akthar-a min **raqs-at**-in — Arabic
danced DEF-man.PL more-ACC than dance-UNIT-GEN
'*The men danced three times*'
From (Fassi Fehri, 2012, p.309)

Fassi Fehri (2012) does not go into detail regarding whether eventive -AH is related to the physical singulative -AH, but their glossing of the examples implies that event -AH is taken to be unitising in some way. As the Arabic feminine marker -AH is a highly productive marker used outside of singulativisation, and the fact that this use of eventive -AH cuts across grammatical domains (verbal/nominal), I will stay ambiguous as the whether it is related to physical singulative -AH. Like artificial collective singulatives, I will therefore have very little to say about non-physical singulativisation, and I present this data here only for a more holistic picture.

2.1.5 Faux singulatives

I coin the term `FAUX SINGULATIVE` for terms that on the surface seem to be part of the collective/singulative class, but they are not (*c.f.* Acquaviva, 2015). Faux singulatives are abundant in Welsh, exemplified in Table 2.8. For these terms, the singular form (sometimes optionally) ends in -EN/-YN. Descriptively, the plural is formed via *replacing* the singular ending with a plural ending. This marking pattern essentially makes singular faux singulatives formally indistinguishable from canonical individuating singulatives.

Singular	**Plural**	**Gloss**
cerig-yn	cerig-os	pebbles→ pebbles
rhesin(-en)	rhesin-s	raisin→ raisins
nion-yn	nion-od	onion→ onions
trog-en	trog-od	tick→ ticks
malw-en	malw-od	snail→ snails
gwilth-en	gwilth-od	slug→ slugs
sionc-yn (gwair)	sionc-od (gwair)	grasshopper→ grasshoppers
chwil-en ddu	chwil-od ddu	cockroach→ cockroaches
chwil-en	chwil-od	beetle→ beetles
eog(-yn)	eog-iaid	salmon→ salmons
lled-en	lled-od	flounder→ flounders
cwning-en	cwning-od	rabbit→ rabbits
blod-yn	blod-au	flower→ flowers
rhos-yn	rhos-od	rose→ roses
briall-en	briall	primrose→ primroses

Table 2.8: Welsh faux singulatives

I argue that the terms in Table 2.8 are not part of the collective/singulative class on the following grounds. Firstly, there is no morphosyntactic evidence that the singular term is derived from the plural term. Rather, the pattern seems to be of suppletion where the singular ending and the singulative marker are syncretic. This claim is supported by general plural marking patterns in Welsh. As Jones and Thomas (1977, p160-161) show, and as shown in Table 2.1 in §2.1.1, suppletion of singular -YN/-EN is one of the general plural-forming strategies in Welsh for some terms that are decidedly not in the collective/singulative class but are in the singular/plural class (*e.g.* MATS-EN → MATS-YS (match-SG → match-PL = *'match → matches'*) and CERP-YN → CARP-IAU (rag-SG → rag-PL = *'rag → rags'*). Secondly, the plural terms in Table 2.8 are not (weakly) number neutral, a feature we will see in §2.2 is defining for collective terms. With this, I follow Jones and Thomas (1977) that such pairs should be described as part of the singular/plural class, rather than part of the collective/singulative class.

Though the terms in Table 2.8 are not part of the collective/singulative class, it may be the case that some of the singular terms in Table 2.8 were *once* part of the collective/singulative class and the singulative term was reanalysed as a singular, and subsequently assigned suppletive number marking.[16] One such example of reanalysis is CWNINGEN (*'rabbit'*), which has been ascribed either collective/singulative or faux singulative status in different Welsh dictionaries. *Geiriadur yr Academi*[17] lists the modern Welsh pair as suppletive CWNING-EN → CWNING-OD (rabbit-SG → rabbit-PL = *'rabbit → rabbits'*), making CWNINGEN part of the singular/plural class (*i.e.* a faux singulative). As far as I am aware, this is the pattern typically used by most Welsh speakers in Northern Wales, and is the pattern used by the informants of this thesis. Yet, *Geiriadur Prifysgol Cymru Online*[18] lists a collective/singulative contrast CWNING → CWNING-EN (rabbit.COL → rabbit.COL-SING = *'rabbits → rabbit'*) from their first published version of the dictionary in 1956. Similarly, in the academic literature Stolz (2001) lists the collective/singulative CWNING→ CWNING-EN pair. I will have more to say on inter and crosslinguistic variation of singular/plural and collective/singulative 'coding' in chapter 4, but for now I simply suggest that in the case of CWNING there is an ongoing process of change of grammatical categories, where CWNING had once been (for some speakers) part of a collective/singulative pair, but it is now in the process of re-analysis to a singular, fossilising the -EN suffix. It then stands to reason that similar histories and variations may apply to other faux singulatives in Table 2.8, though I leave this suggestion as a curiosity for further investigation.

The case of faux singulatives must be examined on a case-by-case basis in and across languages. For Welsh, it seems that the appearance of faux singulatives arises due to a syncretic pair of singulative marking and singular suppletion. In other languages, faux singulatives (if they exist) may look different. In Arabic, for example, as singulative marked terms are realised with the feminine suffix -AH, then faux

[16]See chapter 4 on the notional criteria of the collective/singulative nominal class.
[17]Available online: https://geiriaduracademi.org (accessed January 2024).
[18]'cwning, cwing' GPC Online. 2014. University of Wales Centre for Advanced Welsh & Celtic Studies, 2014. http://www.geiriadur.ac.uk (accessed January 2024).

singulatives may well arise with terms that are feminine marked. We have already seen where this may be the case. Specifically, in the domain of packaging singulatives, Jaradat and Jarrah (2022) basically take JUBN-AH to be a faux packaging singulative, as they do not analyse JUBN-AH as a (portioning) singulative, but do concede it looks like one.

2.1.6 The challenge

From the preceding discussion, it is clear that there are are at least two types of singulatives: individuating singulatives and packaging singulatives. Of these, the packaging singulative is easier to analyse. Under any given account, the packaging singulative may well be akin to a standard packaging/portioning function. The nature of the individuating singulative is much harder to capture. The immediate question that arises is:

(17) Is the individuating collective/singulative system
  a. a singular/plural system with reverse morphology, or
  b. a distinct (minor) number system?

The question in (17) is a non-trivial one. If it turns out that the the correct treatment of collective/singulative systems is to treat them as reverse marked systems in a style of (17a), then we would expect no syntactic/semantic difference between collective terms and morphologically marked plurals, and no syntactic/semantic difference between singular and singulative terms, except for the reverse marking. If this criteria is met, then the appropriate analysis of singulative marking would be to treat singulative marking as an exponent of the singular, and the collective as an exponent of plurality. An appropriate approach to such a system may be within the remit of the distributed morphology framework with a focus on identifying the distribution of marking patterns. This is the approach taken by Kouneli (2019) for Kipsigis, and Erschler (2022) for Digor Ossetic.

Alternatively, if the proper treatment of the individuating collective/singulative system is to treat it as a distinct number system in the style of (17b), then by hypothesis the collective/singulative system forms a number class that is neither a singular/plural system nor a canonical substance count/mass system, but a third nominal class. Such a treatment would offer an analysis where a.) the collective is the base upon which a singulative is derived and b.) the plural marker and the collective are syntactically and/or semantically distinct, and so singulative markers may pluralise. If this is the case, then the appropriate treatment of the singulative is as a function which derives singularities from collectives. This second option is the overwhelmingly popular approach in the semantic literature in the lexicalist tradition. Scholars that take this approach to singulatives include Grimm (2012a) for Welsh and Daagare and Kagan (2024); Wągiel and Shlikhutka (2023a) for Slavic. I also place Nurmio (2017, 2023) and King (2003) as adjacent to this category. Although they do not offer a formal semantic analysis of singulatives, they do regard the collective/system system as a distinct grammatical category of Welsh.

While I have presented two options for the analysis of collective/singulative system, they are not necessarily in competition. There is a third option, a way which fuses both (17a) and (17b). Such a route is taken in the syntactic literature by Dali (2020); Dali and Mathieu (2021b); Jaradat and Jarrah (2022); Mathieu (2012a, 2013, 2014) who assume a constructionist syntactic model of number, treating number marking as an invariable part of a number projection. For instance, in the Dali and Mathieu (2021b) rendition, singulative and singular marking are both an exponent of semantic singularity under allosemic assumptions, making the collective/singulative system at its heart a reverse morphology system (in line with (17a)). However, Dali and Mathieu adopt a split DP hypothesis of number features and so treat collectives as syntactically and semantically distinct from plural markers, thereby predicting that collective/singulative systems may interact with plural markers.

Finally, though there are different analyses available to capture collective/singulative systems, it may be the case that for one language a reverse morphology analysis may be appropriate, while for another language a tripartite number system be appropriate. This is explicitly argued by Kouneli (2019) and Erschler (2022) who posit that singulative/singular marking in Kipsigis and Digor Ossetic respectively is inherently different from the singular/singulative marking in Welsh and Arabic. In essence, Kouneli (2019) argues that approach (17a) is an appropriate route to take for Kipsigis, while the approach in (17b) is more appropriate for Welsh.[19]

As it turns out, though there is disagreement regarding whether collective/singulative systems are reverse morphology systems or distinct minor number systems, there is a common theme of analysis: collective/singulative systems are akin to object mass/general number systems. Specifically, collective terms denote uncountable multiplicities, akin to object mass and/or general number terms, while singulative marking is typically taken as indicative of a function of individuation, classification, division, or packaging. In this manner, the singulative function is taken to derive singular count terms from (object) mass terms or root concepts, dependent on theoretical assumptions (Dali and Mathieu (2021b); Mathieu (2012a, 2013, 2014); Ojeda (1992); Grimm (2012b, 2018); Nurmio (2017, 2020, 2023); Wągiel and Shlikhutka (2023a); Zabbal (2002) though *c.f.* Erschler (2022); Kouneli (2019)). I will have more to say about the specific consequences of these approaches in chapter 4, but for now I turn to a descriptive investigation of the collective/singulative systems of Welsh and Hejazi/Modern Standard Arabic.

In the following section, I will show that the collective/singulative distinctions in both Welsh and Hejazi/Modern Standard Arabic are not cases of a simple markedness inversion. Rather, I will take the view that the singulative marker itself is indicative of a classificatory function upon a collective. As such, I will not be following a reverse morphology analysis such as in the Kouneli (2019) or Erschler (2022) systems. Instead, I will treat the collective/singulative system as a distinct minor number system, specifically an number neutral uncountable system, where the collective is number neutral uncountable and the singulative is a classifier.

[19]This state of affairs highlights the fact that the term `SINGULATIVE` is used throughout the literature to simultaneously describe morphologically distinct phenomena.

2.2 The count/mass distinction

The COUNT/MASS DISTINCTION is a field of enquiry that falls under the rubric of COUNTABILITY and has been of particular interest for linguists for a century, with discussion of the connection between perceptual reality and linguistic representation, dating back to at least Jespersen (1924). Throughout the literature, the use of certain terms to describe the count/mass distinction varies, so I lay out here my intended use of terms COUNT and MASS.

My use of the term COUNT is fairly uncontroversial, and is taken to refer to terms that are countable (may combine with numerals). The term MASS is intended as a grammatical term which covers non-count predication. In this way, as the concept of an object is abstracted from fundamental reality (what is), so too is the count/mass abstracted from the perceptual reality of the substance/object distinction. It is by now common knowledge that the count/mass distinction is grammatical in nature and not a reflection of the (fundamental or perceived) structure of matter. As such, the term MASS should be read as synonymous with UNCOUNTABLE.[20] In the following discussion, in the case that a mass term denotes objects to which we have pre-existing commitments, I will favour the term OBJECT MASS for the English phenomenon and GENERAL NUMBER for the Mandarin Chinese phenomenon, but I assume they are essentially the same, *i.e.* NUMBER NEUTRAL UNCOUNTABLE. I use different terms for ease of comparative discussion only.[21] In the case that the individuals denoted by a mass term are substances (such as liquid or naïvely homogeneous material), I will use the term STUFF MASS regardless of language.

[20] The choice of the term MASS is due to its common usage in the literature (Quine, 1960; Chierchia, 1998a, 2010b; Rothstein, 2010; Landman, 2020, *a.o.*).

[21] The terminology and descriptions used for this phenomena varies. By far the most common term is OBJECT MASS (Bale and Barner, 2018; Barner and Snedeker, 2005; Erbach et al., 2017, 2019; Inagaki and Barner, 2009; Moltmann, 2021; Rothstein, 2017; Sutton and Filip, 2017; Tsoulas, 2009). Other terms used for this phenomena include FAKE MASS (Alexiadou, 2015; Chierchia, 2010b, 2015, 2021; Kiss et al., 2021), NEAT MASS (Landman, 2020), NATURALLY ATOMIC MASS (Rothstein, 2010), and COLLECTIVE MASS (Bunt, 1985; de Belder, 2013).

There is no shortage of papers defining and redefining tests for countability. For Allan (1980), combination with unit/plural determiners and the status of number agreement (amongst other features) serve as tests of countability. Alternatively, Chierchia (1998a,b, 2010b) focuses on combination with numerals and the presence of plural marking as indicative of count/mass status. The forthcoming investigation will limit discussion of the count/mass distinction to the defining properties listed in (18). The reasoning for this limitation is that while languages exhibit variation with how the count/mass distinction manifests (Chierchia, 1998a, p.57), the properties in (18) are attested in Welsh, Arabic, English and Mandarin Chinese, allowing for crosslinguistic comparison.[22]

(18) `PROPERTIES OF THE COUNT/MASS DISTINCTION`

a. `BASIC MEANINGS:`
   Singular count terms refer to (contextual) singularities, mass terms are cumulative.

b. `COMBINATION WITH NUMERALS:`
   Count terms can combine with numerals, mass nouns cannot combine with numerals.

c. `CLASSIFIERS AND MEASURES:`
   Mass terms use intermediary structure to combine with numerals.

d. `PRE-EXISTING PERCEPTUAL REALITY:`
   Pre-existing commitments to perceptual reality (*i.e.* stuff v object) is preserved.
   (i) Object mass terms (and not substance mass terms) are compatible with `STUBBORNLY DISTRIBUTIVE PREDICATES (STUBS)`.
   (ii) Object mass terms measure by numerosity in `COUNT COMPARISON CONSTRUCTIONS` while substance mass terms measure by volume.

[22] See Allan (1980); Chierchia (1998a); Gillon (1992) for other ‘tests’ and sensitive environments for the count/mass distinction.

The optionality of features in (18) lead to crosslinguistic variation. It emerges that all nouns in Mandarin Chinese are grammatically mass, regardless of whether they denote substances or objects (Krifka, 1995; Chierchia, 1998b, 2010b; Landman, 2020). This is not to say that all nouns in Mandarin Chinese are *literally* mass denoting in any ontological sense, rather Mandarin Chinese has only an object mass/stuff mass distinction (= a general number/stuff mass distinction). English, on the other hand, has a three way count/stuff mass/object mass distinction, where object denoting terms may be grammatically count or grammatically object mass. Finally, as we will soon see, Welsh and Hejazi/Modern Standard Arabic have typical count/stuff mass/object mass distinction, similar to English, where collective terms are akin to object mass terms (and therefore also akin to general number terms).

In the following investigation, two points should be made explicit regarding the data. First, nominal types are restricted, where possible, to terms that denote natural kinds, due to the fact that the Welsh and Hejazi/Modern Standard Arabic collective/singulative class overwhelmingly covers the notional classes of fruits, vegetables, insects and animals. As the English object mass class is typically restricted to grains and artificial collection nouns, English examples are presented using FURNITURE and FRUIT as representatives of the object mass class.

Secondly, the examples presented in this thesis are intended to capture existential readings. Some of the Mandarin Chinese examples in the forthcoming discussion may have interpretations other than the existential readings discussed, namely definite and kind/generic readings (See *e.g.* Cheng and Sybesma, 1999; Yang, 2001, for discussion). Non-existential readings will largely be ignored for the purpose of the forthcoming discussion, and will not be included in the example glosses.

2.2.1 Cumulativity and number neutrality

Count terms are taken to be semantically singular, often modelled as denoting a set of atomic/disjoint individuals. Conversely, mass terms have the property of being `CUMULATIVE` (Chierchia (1998a,b, 2010b, 2021); Krifka (1989); Pelletier (1979);

Rothstein (2021); Quine (1960), though *c.f.* Wierzbicka (1988, p.511)). Cumulativity is where the denotation of a term includes unlimited upwards closure, as defined in (19). Cumulativity should be taken as a linguistic property regarding the denotation of a particular term, and not as a property of the referents of the term (see Rothstein, 2010, for discussions of semantic versus perceptual cumulativity). In other words, in the case that a predicate may refer to either (contextual) singularities or sums, it is cumulative.[23]

(19) `CUMULATIVITY`:
P is cumulative iff: $\forall x \forall y[x \in P \wedge y \in P \rightarrow x \sqcup y \in P]$
P is a cumulative predicate if when x and y are in P, then the sum of x and y is also in P
From Rothstein (2010, p.10)

Cumulativity is not the same as `NUMBER NEUTRALITY`. I follow Kang (1994); Corbett (2000); Rullman and You (2006); Wilhelm (2008); Bale et al. (2010) and Bale and Khanjian (2014) in taking number neutrality to be the property where a term is indiscriminate to number distinctions.[24] That is, a number neutral term may denote either a singularity or a sum of individuals, relative to a property. In this way, number neutrality entails cumulativity, but cumulativity does not entail number neutrality.

(20) `NUMBER NEUTRALITY:`
P is number neutral iff: P may denote either integrated wholes relative to a property, or sums of integrated wholes, relative to a property.

Substance Mass

Cumulativity holds for substance mass predicates such as English WATER, and equivalents in Welsh, Mandarin Chinese and Hejazi/Modern Standard Arabic.

[23] $\sqcup$ is defined in chapter 3.
[24] These authors use the term `GENERAL NUMBER` to capture the same phenomena.

This is shown in examples (21)-(24), where the sentence is defined no matter the amount of water purchased. That is, the sentences are perfectly valid whether 500 millilitres of water were bought or whether one litre of water was bought. As such, all (contextual) individuals of water and their sums are in the extension of WATER/SHUǏ/DDŴR/MUUYAA/MA'.

(21) Edward bought **water** English

(22) Shīyáng mǎi-le **shuǐ** M. Chinese
Shīyáng buy-PST water
'*Shīyáng bought water*' (=any amount)

(23) Prynodd Owain **ddŵr** Welsh
buy.PST.3 Owain water
'*Owain bought Water*' (=any amount)

(24) Malik 'ishtaraa **muuyaa** H. Arabic
Malik buy.PST.M.3 water
'*Malik bought water*' (=any amount)

(25) 'ishtaraa Malik-u **ma'**-an MSA
buy.PST.M.3 Malik-NOM water-ACC
'*Malik bought water*' (=any amount)

Object mass, general number and collectives

Cumulativity and number neutrality are defined for English object mass terms. For example, in (26), cumulativity is shown by the fact that if two individuals, a *chair* and a *table*, are in the extension of FURNITURE, then so is their sum, *chair*$\sqcup$*table*. Number neutrality is shown by the fact that *e.g.* FURNITURE in (26) may legally refer to a single item of furniture. In other words, (26) is defined if Edward bought a chair, or if Edward bought a chair and a table.

(26) Edward saw **furniture/fruit** in the garden English
= any amount of furniture/fruit

Cumulativity and number neutrality has been argued to apply to *all* object denoting nominals in Mandarin Chinese (Krifka, 1995; Chierchia, 2010b), regardless of perceptual status of individuals denoted, hence the term `GENERAL NUMBER`. This behaviour is shown in (27) where NǙHÁI (*'girl'*) is compatible with Shīyáng having seen one, or many, girls in the garden.

(27) Shīyáng zài huāyuàn lǐ kàndào-le **nǚhái** M. Chinese
Shīyáng at garden in see-PST girl
'*Shīyáng saw a girl / girls in the garden*'

Collectives in Welsh and Hejazi/Modern Standard Arabic have the same basic interpretations as object mass and general number terms: cumulativity and number neutrality are defined, as collectives denote (sums of) pre-theoretical wholes, allowing both a single witness and sums of witnesses to be true of them.[25] Note, while number neutrality is possible for collective terms, a sense of plurality is preferred; while the examples in (28)-(30) are possible with singular interpretations, a singulative construction would be more natural in singular contexts. The preference for a multiplicity interpretation of the collective is in line with Wright (1896)'s description of the Arabic collective as referring to a genus, and also King (2003)'s description of the Welsh collective as having a strong sense of being part of a clustered group. It is my intuition that this reading is also in-line with English object mass terms such as FURNITURE and FRUIT. That is, while the predicate may but true of singularities, it is more naturally interpreted as referring to multiplicities.

[25]The number neutrality of collectives is noted by others in the literature, but under different terminology and assumptions. For example, Fassi Fehri (2012, 2004) favour calling the collective 'general number' and 'kind denoting'.

(28) Gwelodd Owain **hwyaid** yn yr ardd — Welsh
see.PST.3 Owain duck in DEF garden
'*Owain saw ducks/a duck in the garden*'

(29) Malik shaaf **baṭ** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 duck in DEF-garden
'*Malik saw a duck / ducks in the garden*'

(30) ra'aa Malik-u **baṭ**-an fi al-hadiqat-i — MSA
see.PST.M.3 Malik-NOM duck-ACC in DEF-garden-GEN
'*Malik saw a duck / ducks in the garden*'

Singular count and singulatives

Neither cumulativity nor number neutrality are defined for singular count terms in English, Welsh Hejazi Arabic, nor MSA. If Edward/Owain/Malik saw two girls in the garden, then the examples in (31)-(34) are not defined. Singular count terms in their uninflected form are strictly singular in existential contexts, *i.e.*, they denote individuals that are (contextually) bounded across time and space.[26] Note, Mandarin Chinese has no singular count terms, so no comparison examples are shown in this subsection.[27]

(31) Edward saw a **girl** in the garden — English

(32) Gwelodd Owain **ferch** yn yr ardd — Welsh
see.PST.3 Owain girl in DEF garden
'*Owain saw a girl in the garden*'

[26]Though *c.f.* Rothstein (2010) who explicitly argues that there are count nominals that are not bounded across time and space in *e.g.* English such as FENCE, WALL and BOUQUET.

[27]Terminology and theoretical assumptions differ in the literature. Regardless, the fact that Mandarin Chinese object-denoting terms are cumulative is never rejected. One example where this is the case is in Zhang (2013)'s book, where number neutrality is defined for Mandarin Chinese object denoting nouns, but Zhang does not use the term `MASS` to describe them. Similarly, some authors such as Corbett (2000) favour the term `GENERAL NUMBER` for terms in *e.g.* Mandarin Chinese, but the same overall effect is seen: terms are number neutral and/or object mass-like.

(33) Malik shaaf **bint** fi al-hadiqa — H. Arabic
Malik see.pst.m.3 girl in def-garden
'*Malik saw a girl in the garden*'

(34) ra'aa Malik-u **fataat**-an fi al-hadiqat-i — MSA
see.pst.m.3 Malik-nom girl-acc in def-garden-gen
'*Malik saw a girl in the garden*'

Singulative marked terms mirror singular count terms in basic interpretations. The individuating singulative is interpreted as denoting singular bound individuals, as seen in (35)-(37). There is no interpretation of cumulativity or number neutrality. The only difference between singular and singulative marked terms is therefore morphological in nature. Explicitly, while singulars gain their strictly singular meanings via no overt marking, singulatives must be overtly marked.

(35) Gwelodd Owain **hwyad-en** yn yr ardd — Welsh
see.pst.3 Owain duck-sing in def garden
'*Owain saw a duck in the garden*'

(36) Malik shaaf **baṭ-ah** fi al-hadiqa — H. Arabic
Malik see.pst.m.3 duck-sing in def-garden
'*Malik saw a duck in the garden*'

(37) ra'aa Malik-u **baṭ-at**-an fi al-hadiqat-i — MSA
see.pst.m.3 Malik-nom duck-sing-acc in def-garden-gen
'*Malik saw a duck in the garden*'

Packaging singulatives are also interpreted as referring to singular wholes, as in (3), repeated for convenience. The difference is that the singular whole is not pre-existing, but derived in context.

(3) Bwytodd Owain **gos-yn** ddoe — Welsh
eat.pst.3 Owain cheese-psing yesterday
'*Owain ate a piece of cheese yesterday*'

Plural marked terms

For completion, note that plural marked terms in English, Welsh and Hejazi/Modern Standard Arabic are defined for cumulativity, but not number neutrality, in upwards entailing contexts. The examples in (38)-(41) are defined only if two or more girls were seen in the garden. I will return to plural marked terms in more detail in §2.3, but for now it is sufficient to note that collective terms and plural marked terms differ with respect to their number neutrality status in that only collectives are number neutral in upwards entailing existential contexts.

(38) Edward saw **girls** in the garden — English

(39) — Welsh

Gwelodd	Owain	**ferch-ed**	yn	yr	ardd
see.PST.3	Owain	girl-PL	in	DEF	garden

‘*Owain saw girls in the garden*’

(40) — H. Arabic

Malik	shaaf	**ban-aat**	fi	al-hadiqa
Malik	see.PST.M.3	girl-PL	in	DEF-garden

‘*Malik saw girls in the garden*’

(41) — MSA

ra’aa	Malik-u	**fatay-aat**-in	fi	al-hadiqat-i
see.PST.M.3	Malik-NOM	girl-PL-GEN	in	DEF-garden-GEN

‘*Malik saw girls in the garden*’

Summary of cumulativity and number neutrality

Table 2.9 summarises the basic interpretations of stuff mass, object mass, general number, singular count and singulative terms in English, Welsh, Hejazi/Modern Standard Arabic, and Mandarin Chinese. The data shows that collective/singulative systems align with object mass and general number systems in their basic interpretations. Singulative marked and singular terms share the same basic interpretations of semantic singularity, while object mass, collective and general number share the same basic interpretations of cumulativity and number neutrality.

	Cumulative	**Number Neutral**	**Semantically Singular**
Substance Mass	✔		
Object Mass	✔	✔	
General Number	✔	✔	
Collectives	✔	✔	
Plural Marked	✔		
Singular Count			✔
Singulative			✔

Table 2.9: Summary of basic meanings for predicate types

2.2.2 Combination with numerals

Count terms are so-called because they may freely combine with numerals. Mass terms cannot freely combine with numerals, at least not with the interpretation of counting discrete individuals (Borer, 2005a; Cheng and Sybesma, 1998; Chierchia, 1998a,b, 2010b, 2021; Doetjes, 2017, 2021; Erbach, 2021; Gillon, 1992; Gleason, 1965; Kiss et al., 2021; Krifka, 1989, 1995; Landman, 2020; Moltmann, 2020, 2021; Pelletier, 1975, 2012; Quine, 1960; Rothstein, 2009, 2010, 2011, 2021, *a.o.*). This property is so robust that that Chierchia (1998a, 2010b, 2021) dubs it the `SIGNATURE PROPERTY`, suggesting it is one of the few crosslinguistic universals with respect to the countability distinctions.

Singular count and singulatives

English, Welsh and Hejazi/Modern Standard Arabic count terms easily combine with numerals, though there is variation with how plurality is marked in numeral constructions. In English, there is compulsory plural agreement with any numeral above two. In Welsh, there is compulsory singular agreement in all cases. In Hejazi/Modern Standard Arabic there is a mixed system, where plural agreement is only seen with numerals under ten. Above ten, the nominal exhibits singular agreement.

(42) Edward saw **five girls** in the garden — English

(43) Gwelodd Owain **bum merch** yn yr ardd. — Welsh
see.PST.3 Owain five girl in DEF garden
'*Owain saw five girls in the garden*'

(44) Malik shaaf **khams-ah ban-aat** fi al-hadiqa — H. Arabic
see.PST.M.3 Malik five-F girl-PL in DEF-garden
'*Malik saw five girls in the garden*'

(45) ra'aa Malik-u **khams**-a **fatay-aat**-in fi al-hadiqat-i — MSA
see.PST.M.3 Malik-NOM five-ACC girl-PL-GEN in DEF-garden-GEN
'*Malik saw five girls in the garden*'

(46) Malik shaaf **khamstashr**-ah **bint** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 fifteen-F girl in DEF-garden
'*Malik saw fifteen girls in the garden*'

(47) ra'aa Malik-u **khams-a 3shr**-at-a **fatah**-an — MSA
see.PST.M.3 Malik-NOM five-ACC ten-F-ACC girl-ACC
fi al-hadiqat-i
in DEF-garden-GEN
'*Malik saw fifteen girls in the garden*'

Crosslinguistic variation in plural marking has spawned a myriad of theoretical work and assumptions regarding the role of plural markers. Some camps of thought include the claims that plural marking is simply agreement (Ionin and Matushansky, 2004, 2006; Borer, 2005a; Wągiel, 2021b), that plural marking provides semantic content (Landman, 2020; Grimm, 2012b), that plural marking is conditioned by output constraints (Farkas and de Swart, 2010), and that plural marking is a 'check point' for atomicity (Chierchia, 2010b). Regardless of the theoretical backdrop that underlies opinions, the `SIGNATURE PROPERTY` (*i.e.* whether numerals and nouns combine) is independent of whether a language experiences obligatory pluralisation with numeral.[28] This is an enlightening view, as it turns out that singulative marked terms pattern exactly as singular terms with respect to their numeral agreement patterns, indicating that singulative and singular terms undergo the same processes to combine with numerals. The case of Welsh is straightforward: there is obligatory non-pluralisation of the singulative.[29]

(48) Welsh

Gwelodd	Owain	**bum**	**hwyad-en**	yn	yr	ardd
see.PST.3	Owain	five	duck-SING	in	DEF	garden

‘*Owain saw five ducks in the garden*’

The case of Hejazi/Modern Standard Arabic is more intricate, but fundamentally singulative marked terms behave as singulars with respect to their behaviour when combining with numerals.[30] There is compulsory pluralisation of the singulative when it occurs with numerals below ten and compulsory singular agreement with numerals higher than ten. For all terms in the collective/singulative class, the compulsory agreement with numerals under ten may be realised by the feminine sound plural form (-AAT). In these cases, the singulative marker -AH is absorbed via general lengthening patterns (see §2.1.1).

(49) H. Arabic

Malik	shaaf	**khams**-ah	**baṭ-a-at**	fi	al-hadiqa
Malik	see.PST.M.3	five-F	duck-SING-PL	in	DEF-garden

‘*Malik saw five ducks in the garden*’

(50) MSA

ra’aa	Malik-u	**khams**-a	**baṭ-a-at**-in
see.PST.M.3	Malik-NOM	five-ACC	duck-SING.PL-GEN

fi	al-hadiqat-i
in	DEF-garden-GEN

‘*Malik saw five ducks in the garden*’

[28] This claim is also made by Chierchia (2010b) who uses Finnish as the parade example of why plural marking in numeral contexts and the count/mass distinction are independent. Rothstein (2009, p.108, fn) makes a similar claim for Turkish. Additionally, Wilhelm (2008) shows that there is a count/mass distinction in Dëne Sųłiné, but there is no plural marking on count terms.

[29] The Welsh facts are also pointed out by Roberts and Gathercole (2012, p.71).

[30] The facts for Arabic are also seen for Maltese singulatives. See Corbett (1996, p.11) for more details.

(51) Malik shaaf **khamstashr**-ah **baṭ-ah** fi al-hadiqa H. Arabic
see.PST.M.3 Malik fifteen-F duck-SING in DEF-garden
'*Malik saw fifteen ducks in the garden*'

(52) ra'aa Malik-u **khams**-a **ashr**-at-a **baṭ-at**-an fi al-hadiqat-i MSA
see.PST.M.3 Malik five-ACC ten-F-ACC duck-SING-ACC in DEF-garden-GEN
'*Malik saw fifteen ducks in the garden*'

For some terms in the collective/singulative class, the plural of the singulative form may be realised not only as a feminine sound plural, but also a broken plural, where stem changes absorb singulative morphology as in (53) for Hejazi Arabic and (54) for Modern Standard Arabic, where the a examples are the feminine sound plural of the singulative and the b examples are the broken plural of the singulative.[31]

(53) a. Malik shaaf **khams**-ah **shajar-a-at** fi al-hadiqa H. Arabic
Malik see.PST.M.3 five-F tree-SING-PL in DEF-garden
'*Malik saw five trees in the garden*' = (53b)

b. Malik shaaf **khams**-ah **'ashjaar** fi al-hadiqa H. Arabic
see.PST.M.3 Malik five-F tree.SING.PL in DEF-garden
'*Malik saw five trees in the garden*' = (53a)

[31]Though singulative marking is not visible in plural contexts, it is nonetheless present, as shown by agreement patterns. In Arabic, collectives are masculine and singulatives induce gender shift to feminine (Fassi Fehri, 2004, 2012, 2020; Mathieu, 2012a, 2013, 2014, *a.o.*). In MSA, numerals observe reverse agreement with the singular form of the noun, and in Hejazi Arabic numerals agree in gender with the singular form of the noun. The fact that the numeral is masculine in MSA and feminine in Hejazi Arabic indicate that singulativisation has taken place.

(54) a. ra'aa Malik-u **khams**-a **shajar-a-at**-in MSA
see.PST.M.3 Malik-NOM five-ACC tree-SING-PL-GEN
fi al-hadiqat-i
in DEF-garden-GEN
'*Malik saw five trees in the garden*' = (54b)

b. ra'aa Malik-u **khams**-a **'ashjaar**-in MSA
see.PST.M.3 Malik-NOM five-ACC tree.SING.PL-GEN
fi al-hadiqat-i
in DEF-garden-GEN
'*Malik saw five trees in the garden*' = (54a)

Not every singulative has a broken plural of the singulative form. While SHAJAR-AH (tree.COL-SING = *'tree'*), SAMAK-AH (fish.COL-SING = *'a fish'*) and TUM-AH (date.COL-SING = *'date'*) have broken plurals, there are no broken plural forms for BAṬ-AH (duck.COL-SING = *'duck'*), DAJAAJ-AH (chicken.COL-SING = *'chicken'*) and BURTAGAAL-AH (orange.COL-SING = *'orange'*). In the case that a singulative has both feminine sound plural and broken plural forms, there is no interpretative difference between the two in numerical contexts.[32]

Substance Mass

Combining numerals and substance mass terms is extremely marked; it is either outright ungrammatical or requires reinterpretation to contextualised singulars. In English, [number + stuff mass] combinations may only be interpreted as contextually portioned amounts. In the literature, the coerced portion interpretation is often referred to as `PACKAGING` (Pelletier, 1975/1979; Bach,1986a; Jackendoff,1991), and the sub-kind interpretation as `SORTING` (Bunt, 1985).

[32] That some singulative forms have no broken plural forms is a matter of morphological realisation constraints. See McCarthy and Prince (1990) for discussions of the morphological conditions that must be met for a broken plural to arise (*e.g.* moras, number of syllables).

(55) Edward bought five waters — English
= Edward bought five portions of water (packaging)
= Edward bought five kinds of water (sorting)

As pointed out by Doetjes (2021, p.112), there is crosslinguistic variation in the availability of coerced readings. Doetjes states that while TWO GOLDS is perfectly adequate in English in the context of Olympic gold medals, the parallel construction is not available in Dutch. Returning to our languages of interest, similar observations hold, where combining stuff mass directly with numerals has different effects for each language. In Welsh, for some speakers there is a marginal packaging and sorting coercion for [number + stuff mass] constructions, but most often these structures are rejected. Similar observations hold for Hejazi/Modern Standard Arabic, where coercion of substance mass in numeral contexts is marginal.[33] In Mandarin Chinese, the picture is more extreme. All [number + stuff mass] combinations are outright ungrammatical. No packaging or sorting interpretation may arise in these contexts.[34]

(56) #Prynodd Owain **bum dŵr** — Welsh
buy.PST.3 Owain five water
Marginally: '*Owain bought types/portions of water* '

(57) #Malik 'ishtaraa **khams**-ah **muuyaa-at** — H. Arabic
Malik buy.PST.M.3 five-F water-PL
Intended: '*Malik bought portions/types of water*'

(58) #'ishtaraa Malik-u **khams**-a **miyyaah**-in — MSA
buy.PST.M.3 Malik-NOM five-ACC water.PL-GEN
Intended: '*Malik bought types/portions of water*'

(59) *Shīyáng mǎi-le **wǔ shuǐ** — M. Chinese
Shīyáng buy.PST five water
Intended: '*Shīyáng bought types/portions of water*'

[33] *c.f.* Fassi Fehri (2012) who holds coercion is attested in Arabic.
[34] *c.f.* Zhang (2013, p.23) for contexts where packaging may be indirectly attested.

Object mass and collectives

At first it may seem that the non-acceptability of [number + stuff mass] terms is due to substance mass terms not providing pre-theoretical individuals which may be counted. Yet, it is continuously shown that the non-acceptability of combining numerals with mass terms is not simply due to structure of matter (Chierchia, 1998a, 2010b; Krifka, 1989; Rothstein, 2010, 2021; Landman, 2020, *a.o.*). As is well known, object mass terms such as FURNITURE or FRUIT in English cannot directly combine with numerals with the intention of counting discrete individuals. Instead, the interpretation of (60), if any, is coerced. There is a marginal reading where Edward may have see five kinds of furniture in the garden, or five kinds of fruit in the bowl.[35]

(60) a. #Edward saw **five furniture(s)** in the garden — English
= Edward saw five kinds of furniture in the garden (sorting)
b. #There's **five fruit(s)** in the bowl
= Edward saw five kinds of fruit in the bowl (sorting)

Similarly, in Mandarin Chinese, no general number term may directly combine with numerals (Krifka, 1995; Chierchia, 2010b; Rothstein, 2010; Cheng and Sybesma, 1998). Explicitly, (61) is outright ungrammatical, as are all nouns that appear with a numeral sans classifier.

(61) *Shīyáng zài huāyuán lǐ kàndào-le **wǔ nǚhái** — M. Chinese
Shīyáng at garden in see.PST five girl
'*Shīyáng saw five girls in the garden.*'

Collective terms in Welsh and Hejazi/Modern Standard Arabic observe similar behaviour: they do not directly combine with numerals with the intention of counting discreet entities. In Hejazi/Modern Standard Arabic, the combination of a numeral

[35] *c.f.* Cowper and Hall (2012) who claim the universal sorter is not available for object mass terms such as FURNITURE.

and a collective, as in (63) and (64), results in felicitous interpretations only through a coerced reading.[36] In Welsh, coerced readings are not attested (62).

(62) *Gwelodd Owain **bump hwyaid** yn yr ardd — Welsh
see.PST.3 Owain five duck in DEF garden
Intended: '*Owain saw five types of duck in the garden*'

(63) #Malik shaaf **khams baṭ** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 five duck in DEF-garden-GEN
Intended: '*Malik saw five kinds of duck in the garden*'

(64) #ra'aa Malik-u **khams-at-a baṭ**-in fi al-hadiqati — MSA
see.PST.masc Malik-NOM five-F-ACC duck-GEN in DEF-garden-GEN
Intended:'*Malik saw five kinds of duck in the garden*'

While this thesis will not delve into (the absence of) sorter or packager interpretations in and across languages, such observations show that the substance/object distinction is independent of the countability of object mass, collective and general number terms. This is because while object mass, collective and general number terms denote discrete individuals, they nonetheless do not directly combine with numerals to count those individuals.

Summary of combination with numerals

Table 2.10 summarises the grammaticality of combining singular count, plural count, stuff mass, and object mass, collective and general number terms with numerals. The data shows that count terms and singulatives are aligned in their ability to combine with numerals, including singulatives and singulars experiencing identical language specific agreement patterns. This indicates that the function of the singulative is to derive a singular count term, and then, as a singular count term, the singulative is subject to the same processes that apply to unmarked singular count terms. Also

[36]Similar observations are captured in Fassi Fehri (2012, 2020); Jaradat and Jarrah (2022) and Ouwayda (2014) for Arabic.

aligned is the umbrella category of mass terms, which cannot directly combine with numerals including stuff mass, but also object mass, collective and general number.

	Can combine with numerals?	**Agreement pattern**	
Substance Mass	No		-
Object Mass	No		-
General Number	No		-
Collectives	No		-
Singular count	Yes	English	Compulsory pluralisation
		Arabic	Compulsory pluralisation below 10; singular agreement above 10
		Welsh	Compulsory Singular agreement
Singulative	Yes	Arabic	Compulsory pluralisation below 10; singular agreement above 10
		Welsh	Compulsory Singular agreement

Table 2.10: Summary of combination with numerals

2.2.3 Measures and classifiers

Combining grammatically mass terms with numerals typically requires a mediator, such as a `CONTAINER PHRASE`, `MEASURE PHRASE`, `PARTITION`, or a `CLASSIFIER` where the functions of these intermediatory phrases derive a grammatical count term from a grammatical mass term (Chierchia, 1998a, 2010b, 1998b; Cheng and Sybesma, 1998; Doetjes, 2017; Kiss et al., 2021; Krifka, 1989, 1995; Moltmann, 2020, 2021; Rothstein, 2010, 2011, 2017, *a.o.*).[37]

[37] Different types of classifiers have different names in the literature. Cheng and Sybesma (1998) use the term `COUNT CLASSIFIER` for classifiers that name the natural unit, and `MASSIFIER` for classifiers which creates individuals in context. In other work, Cheng et al. (2012) use the terms `SORTAL CLASSIFIERS` and `NON-SORTAL CLASSIFIERS` for a similar distinction.

Classifiers and measures differ in an important aspect with regard to how singularities are modelled with respect to perceptual realities. Container phrases, measure phrases and partitions impart some (contextual) standard upon measurement, while classifiers do not. Put succinctly by Cheng and Sybesma (1998), measures *create* individuals which we count, but classifiers *name* pre-existing individuals to which we count.

A further important notion is that some classifiers may have perception-based restrictions such that they may only combine with either object mass or stuff mass terms. For example, the Mandarin Chinese general classifier GÈ (to be discussed below) is often said to only combine with object denoting terms (Chao, 1968; Cheng and Sybesma, 1998; Chierchia, 2021). Due to this, some authors suggest that the count/mass distinction in generalised classifier languages is at the level of the classifier, rather than at the level of the nominal term.[38]

Containers, measures and partitions

We start first with the CONTAINER PHRASE. There are two subtypes of container phrases: those which have CONTAINER readings, and those which have CONTENTS readings (Rothstein, 2010; Khrizman et al., 2015; Cheng and Sybesma, 1998; Cheng, 2012). The readings are made explicit by the comparison of the a and b examples in (65)-(69). In the a sentences, the bottle is interpreted as part of what was purchased.

[38]This specific use of terminology in the literature is, I believe, confusing. The discussion of the count/mass distinction at the classifier level in Mandarin Chinese relies on perceptual realities, such as objecthood and substancehood and not grammatical status/category. Indeed, Chierchia (2021) casts the discussion of classifiers and the count/mass distinction in terms of whether they are COGNITIVELY COUNT, a clearly perceptual, and not linguistic, concept. It seems to me that the difference between the notional restriction of classifiers is perceptual in nature. I suggest that the way this interacts with the grammar is as follows: in Mandarin Chinese there is a selectional requirement for (some) classifiers, such that they may combine with wither a stuff mass or general number (number neutral uncountable) term. As stuff mass terms always denote substances, and general number terms objects, this gives rise the a quasi-perceptional requirement for legal classifier combinations. With this, I avoid using count/mass terminology at the classifier level, and stick to using COUNT as a grammatical term. As such, I do no deny the phenomena that classifiers have selectional requirements that rely on pre-existing perceptual commitments, but I do not call this a count/mass distinction.

This is the container reading, where the container itself is part of the individual denoted. In the b sentences, the bottle is not part of what was drunk. This is the contents reading, where the container is not considered part of the relevant measured individual.[39]

(65) a. Edward bought five **bottles** of water. English

b. Edward drank five **bottles** of water. English

(66) a. Shīyáng mǎi-le wǔ **píng** shuǐ M. Chinese
Shīyáng buy-PST five bottle water
'*Shīyáng bought five bottles of water*'

b. Shīyáng hē-le wǔ **píng** shuǐ M. Chinese
Shīyáng drink-PST five bottle water
'*Shīyáng drank five bottles-worth of water*'

(67) a. Prynodd Owain bum **potel** o ddŵr Welsh
buy.PST.3 Owain five bottle of water
'*Owain bought five bottles of water*'

b. Yfodd Owain bum **potel** o ddŵr Welsh
drink.PST.3 Owain five bottle of water
'*Owain drank five bottles-worth of water*'

(68) a. Malik 'ishtaraa khams-at **3lb** muuyaa H. Arabic
Malik buy.PST.M.3 five-F bottle.PL water
'*Malik bought five bottles of water*'

b. Malik shariba khams-at **3lb** muuyaa H. Arabic
Malik drink.PST.M.3 five-F bottle.PL water
'*Malik drank five bottles-worth of water*'

[39] Similar observations are discussed by de Vries and Tsoulas (2021) who note the incorporation of packaging material into interpretations of the noun.

(69) a. 'ishtaraa Malik-u khams-a **zujaja-at**-in min al-ma' MSA
buy.PST.M.3 Malik-NOM five-ACC bottle-PL-GEN of DEF-water
'*Malik bought five bottles of water*'

b. shariba Malik-u khams-a **zujaja-at**-in min al-ma' MSA
drink.PST.M.3 Malik-NOM five-ACC bottle-PL-GEN of DEF-water
'*Malik drank five bottles-worth of water*'

Similar to container readings, there are `MEASURE PHRASES`, so-called as they measure precise amounts. Like container phrases, the standard of individuation which is counted is built into the measure phrase. Examples of measure phrases are given in (70)-(74), where the amount of water purchased equals five litres.

(70) Edward bought five **litres** of water English

(71) Prynodd Owain bum **litr** o ddŵr Welsh
buy.PST.3 Owain five litre of water
'*Owain bought five litres of water*'

(72) Shīyáng mǎi-le wǔ **shēng** shuǐ M. Chinese
Shīyáng buy-PST five litre water
'*Shīyáng bought five litres of water*'

(73) Malik 'ishtaraa khams-at **litr-aat** muuyaa H. Arabic
Malik buy.PST five-F litre-PL water
'*Malik bought five litres of water*'

(74) 'ishtaraa Malik-u khams-a **litr-aat** min al ma' MSA
buy.PST.M.3 Malik-NOM five-ACC litre-PL of DEF water
'*Malik bought five litres of water*'

A final type of measure-based phrase is that which Chierchia (2010b, 2017) terms the `QUANTITY MEASURE` which does not measure precise amounts. According to Chierchia (2010b), the term QUANTITY in English is a relational vacuous counting phrase. Chierchia argues that whatever is denoted by 'QUANTITY OF WATER' is also denoted by WATER, and as such QUANTITY incurs no special meaning, and is simply used to create (contextually defined) partitions for counting purposes. This is shown in (75a), where the amount of water drunk may be vastly different in different contexts. In a footnote, Chierchia (2010b, p.24) extends the analysis of QUANTITY to English PIECE on stuff mass terms, where a simple partition with no commitment to size/shape is created, as in (75b). In the full analysis, Chierchia suggests that PIECE is ambiguous between a `QUANTITY` reading and a `PART` reading, *e.g.* PIECE in (76) is interpreted as a part-of relation, while PIECE in (75b) is interpreted as a quantity phrase.

(75) Quantity Phrase:
   a. Edward drank five quantities/amounts of water — English
   b. Edward bought five pieces of gold

(76) Edward bought a piece of that pizza — English

The quantity phrase is mirrored crosslinguistically. I suggest that in Mandarin Chinese the quantity phrase appears as a stand alone morph as in (77), and in Welsh, the quantity phrase is the basic interpretative function of the packaging singulative as in (78).

(77) Shīyáng zuótiān chī-le wǔ **kuài** qǐsī — M. Chinese
   Shīyáng yesterday eat-PST five CL.SLICE cheese
   '*Shīyáng ate five pieces of cheese yesterday*'

(78) Bwytodd Owain bum gos-**yn** ddoe — Welsh
   eat.PST.3 Owain five cheese-PSING yesterday
   '*Owain ate five pieces of cheese yesterday*'

Classifiers and individuating singulatives

Unlike measures and partitions, classifiers do not impose any perceptual commitments regarding measurement. Classifiers do not 'create' individuals, but simply name individuals for which perceptual commitments already exist. Only once a classier is present may an object mass or general number term combine with numerals. There are two types of classifiers: `GENERAL CLASSIFIERS` and `SPECIFIC CLASSIFIERS`. General and specific classifiers differ in that general classifiers have no notion of size or shape built into them, but specific classifiers do (Doetjes, 2021). Examples of specific classifiers are English KERNEL in (79), which is used to count grains, and Mandarin Chinese ZHĪ in (80) which is used to count small animals.

(79) Edward ate **five kernels of corn** — English

(80) Shīyáng zài huāyuán lǐ kàndào-le **wǔ zhī yāzǐ** — M. Chinese
Shīyáng at garden in see-PST five CL duck
'*Shīyáng saw five ducks in the garden*'

Examples of general classifiers are seen in the non-part/whole use of English PIECE in (81), and in Mandarin Chinese by the generalised classifier GÈ in (82). In almost all object mass/general number numeral contexts, a general classifier may be used in alternation with the specific classifier (Chao, 1968).[40]

(81) Edward saw **five pieces** of **furniture/fruit** in the garden — English

(82) Shīyáng zài huāyuán lǐ kàndào-le **wǔ gè nǚhái** — M. Chinese
Shīyáng at garden in see-PST five CL girl
'*Shīyáng saw five girls in the garden*'

[40] See Zhang (2013, p46-48) for discussion of some cases where the generalised classifier GÈ cannot replace a more specific classifier.

At this point, recall that collective terms in Welsh and Hejazi/Modern Standard Arabic do not directly combine with numerals. However, as shown in examples (48), (49) and (50), repeated below, singulative marked terms can combine with numerals, and when doing so exhibit the standard morphosyntactic agreement patterns that singular terms exhibit in these environments.

(48) Gwelodd Owain **bum hwyad-en** yn yr ardd — Welsh
see.PST.3 Owain five duck-SING in DEF garden
'*Owain saw five ducks in the garden*'

(49) Malik shaaf **khams**-ah **baṭ-a-at** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 five-F duck-SING-PL in DEF-garden
'*Malik saw five ducks in the garden*'

(50) ra'aa Malik-u **khams**-a **baṭ-a-at**-in — MSA
see.PST.M.3 Malik-NOM five-ACC duck-SING.PL-GEN
fi al-hadiqat-i
in DEF-garden-GEN
'*Malik saw five ducks in the garden*'

As it is the case that individuating singulative marking is required for collectives to combine with numerals, the natural conclusion is that the singulative morpheme is a classificatory function in nature, a fact commonly noted in the literature (Ouwayda, 2014; Fassi Fehri, 2004, 2012, 2016, 2020; Jaradat and Jarrah, 2022; Mathieu, 2013, 2014, 2012a; Zabbal, 2002). While taking the individuating singulative to be classificatory in nature is not controversial, Nurmio (2023) points out one difference between classifiers and singulatives: classifiers are often nouns with extra meaning such as size and shape, while singulatives are suffixes with no meaning other than individuation. While this is true, I argue that this points to an analysis where singulative morphemes are akin to a general classifiers such as English PIECE or Mandarin Chinese GÈ.

Before moving on, note that when classifiers appear bare in English and Mandarin Chinese, their interpretation parallels bare singulatives: strict singularity.[41]

(35) Welsh

Gwelodd	Owain	**hwyad-en**	yn	yr	ardd
see.PST.3	Owain	duck-SING	in	DEF	garden

'*Owain saw a duck in the garden*'

(36) H. Arabic

Malik	shaaf	**baṭ-at**	fi	al-hadiqa
Malik	see.PST.M.3	duck-SING	in	DEF-garden

'*Malik saw a duck in the garden*'

(83) MSA

ra'aa	Malik-u	**baṭ-at**-an	fi	al-hadiqat-i
see.PST.M.3	Malik-NOM	duck-SING-ACC	in	DEF-garden-GEN

'*Malik saw a duck in the garden*'

(84) Edward saw a **piece of furniture** in the garden — English

(85) M. Chinese

Shīyáng	zài	huāyuán	lǐ	kàndào-le	**gè**	**nǚhái**
Shīyáng	at	garden	in	see-PST	CL	girl

'*Shīyáng saw a girl in the garden.*'

Pseudopartitive

The individuating singulative is not the only mediatory structure that collectives may use to combine with numerals. Both Arabic and Welsh may use pseudopartitive structures to combine collectives with numerals. In the case of Arabic, semi-lexical heads are used as in (86).[42] It has been noted by Fassi Fehri (2020, p.86) that the choice between the singulative and the pseudopartitive in Arabic has little interpretative difference.

[41] See also Cheng and Sybesma (1999); Cheng et al. (2012) for discussion of bare classifiers in Mandarin Chinese.

[42] Similarly, Jaradat and Jarrah (2022) report that in Jordanian Arabic the classifier HABB is interchangeable with the singulative.

(86)

khams-at-u	**ru'uus-i**	**(min)**	**baqar**-in	MSA
five-F-NOM	heads-GEN	of	cow-GEN	

'*Five heads of cattle*'

From Fassi Fehri (2020), adapted

For Welsh, a similar story is seen, but in this case there is no semi-lexical head. Nonetheless, there is no interpretative difference between the use of a pseudopartitive and the singulative construction.[43] Note, unlike singulatives the pseudopartitive construction must appear with a numeral.

(87)

Gwelodd	Owain	**bump**	**o**	**hwyaid**	yn	yr	ardd	Welsh
see.PST.3	Owain	five	of	duck	in	DEF	garden	

'*Owain saw five ducks in the garden*' = (48)

(88)

*Gwelodd	Owain	**o**	**hwyaid**	yn	yr	ardd	Welsh
see.PST.3	Owain	of	duck	in	DEF	garden	

Intended: '*Owain saw a duck in the garden*'

Summary of Measures and Classifiers

Table 2.11 summarises the features of measures, classifiers and singulatives. The data shows that the individuating singulative has the same basic classificatory function as general classifiers such as English PIECE and Mandarin Chinese GÈ: a.) they derive singular individuals while maintaining commitments to pre-existing individuals b.) they are required for object mass, collective and general number to combine with numerals. The packaging singulative on the other hand has a partition function in nature, such as English AMOUNT.

[43]The pseudopartitive [num + o + noun] structure is briefly noted by Mittendorf and Sadler (2005) and Borsley et al. (2007, p.170), but their discussion concerns the pseudopartitive on plural structures, not on collective structures. Mittendorf and Sadler reports that there is little semantic difference between the use of a [num + singular-noun] structure and a [num + o + plural-noun] structure, while Borsley et al. reports that the [num + o + noun] structure is used predominantly with higher numerals.

	Combines with...	**Individuals denoted**
Container phrases	stuff mass	Derived via classifier
Measures	stuff mass	Derived via classifier
Partitions	stuff mass	Derived via context
Packaging singulative	stuff mass	Derived via context
Specific Classifiers	restricted object mass/general number	Pre-existing
General Classifier	object mass/general number	Pre-existing
Individuating singulative	collectives	Pre-existing
Pseudopartitive	collectives	Pre-existing

Table 2.11: Summary of measures and classifiers

2.2.4 Pre-existing perceptual reality

The count/mass distinction does not inherently reflect any perceptual distinction regarding the structure of matter. Nevertheless, there are grammatical distinctions between object mass and stuff mass terms that seemingly mirror perceptual realities, *i.e.* object mass terms have commitments to pre-existing individuals, and stuff mass terms do not (Chierchia, 1998a,b, 2010b, 2017, 2021; Rothstein, 2010, 2017, 2021).[44]

We have so far seen reflexes of this in the classifier systems of English and Mandarin Chinese where specific classifiers (such as KERNEL, ZHĪ) may only combine with terms which denote particular objects. Further, general classifiers and singulatives maintain pre-existing commitments of perceptual reality, *i.e.* they

[44]While commitment to objecthood is not typically denied, authors vary on their terminology for this phenomena. For example, Chierchia (2021) uses the term `COGNITIVELY COUNT` and Rothstein (2010) uses the term `NATURAL ATOMICITY` to describe the fact that object mass terms denote individuals that can be counted at a cognitive, but not linguistic level. This discussion concerns levels of ontological modelling, *i.e.* where construed ontologies and linguistic ontologies interact, which I discuss in more detail in chapter 4. I avoid terms such as `COGNITIVELY COUNT` as not to cause any inadvertent confusion with `COUNT` in a grammatical sense. I also avoid `NATURAL ATOMICITY` due to my preference to model integrated wholes as maximally strongly self connected in a mereotopological manner (see chapter 3).

maintain objecthood of the individuals denoted by object mass, collective and general number terms, while measure based classifiers and packaging singulatives combine with stuff mass terms and create individuals in context. In this section, I briefly present two other grammatical reflexes that are sensitive to pre-existing perceptual commitments: `STUBBORNLY DISTRIBUTIVE PREDICATES` and `COMPARISON CONSTRUCTIONS`. In a nutshell, these phenomena each presuppose that nominal terms denote objects rather than substances.

Stubbornly Distributive Predicates (STUBs)

Object mass, collective and general number terms are compatible with what has now come to be known as `STUBBORNLY DISTRIBUTIVE PREDICATES` (`STUBS`), a phenomenon which is often attributed to both Scwarzchild (2011) and Rothstein (2010). Though the term was popularised due to Scwarzchild's work, the basic observation dates back to at least Quine (1960) and McCawley (1975).[45] `STUBS` are terms which combine with and subsequently distribute over each individual in a nominal denotation. Crucially, `STUBS` allow *only* a distributive reading. For example, HEAVY in (89) is not a `STUB`, because while there is a distributive reading where each girl is heavy, there is also a `GROUP` reading where there is no need for any girl to be individually heavy, but what is heavy is the group as a whole. Contrastively, BIG in (90) is a `STUB`, as each girl is individually big. There is no group reading where a group of girls together are taken as big.

(89) The girls are **heavy** English
= The girls are each individually heavy (distributive reading)
= The girls, together, are heavy (group reading).

(90) Those girls are **big** English
= The girls are each individually big (distributive reading)

[45] Quine (1960) was early to notice that 'non-cumulative' adjectives do not occur with mass terms: *SPHERICAL WINE and *SQUARE WATER. This was expanded in McCawley (1975, p.319) who shows that size adjectives are compatible with object mass terms, *e.g.* LARGE FURNITURE.

`STUBS` are sensitive to perceptual realities as they require the nominal they distribute over to denote bounded individuals. As such, `STUBS` are compatible with not only plural terms as in (90), but also object mass terms as in (91). In the case that a nominal has no commitments to bounded individuals, such as a stuff mass term, the use of a `STUB` is ungrammatical, as in (92).

(91) This furniture/fruit is **big** — English
Edward has **big** furniture/fruit — English
= Each piece of furniture/fruit is individually big (distributive reading)

(92) *This water is **big** — English
*Edward has **big** oil — English

The observations in English hold also for Welsh, Hejazi/Modern Standard Arabic and Mandarin Chinese. Nominal terms which denote objects are compatible with `STUBS`, including plural marked (93) and collective/general number terms (94). In neither Welsh, Hejazi/Modern Standard Arabic, nor Mandarin Chinese can `STUBS` combine with stuff mass terms, as shown in (95).[46]

(93) a. Mae'r ferch-ed i gyd yn **fawr** — Welsh
be.3-DEF girl-PL in all YN big
'*Those girls are big*'

b. duul al-ban-aat **kabar** — H. Arabic
those DEF-girl-PL big
'*Those girls are big*'

c. al-fatay-aat-u **kabiir**-a-at-un — MSA
DEF-girl-PL-NOM big-F-PL-NOM
'*The girls are big*'

[46] That Mandarin object denoting nouns, but not substance denoting nouns, may combine with `STUBS` is also explicitly discussed by Zhang (2013), though their terminology is vastly different that the terminology I use here. See also Kang (1994) for discussions of distributive predicates in Korean, another 'mass language'.

(94) a. Mae'r hwyaid i gyd yn **fawr** — Welsh
be.3-DEF duck in all YN big
'*Those ducks are big*'

b. daa al-baṭ **kabar** — H. Arabic
this DEF-duck big
'*Those ducks are big*'

c. hadhaa al-baṭ-u **kabiir**-un — MSA
this DEF-duck-NOM big-NOM
'*Those ducks are big*'

d. Zhè-xiē māo hěn-**dà** — M. Chinese
this-CL.PL cat very-big
'*These cats are big*'

(95) a. *Mae'r ddŵr i gyd yn **fawr** — Welsh
be.3-DEF water in all YN big
'*This water is big*'

b. *duul al-muya **kabar** — H. Arabic
this DEF-water big
'*This water is big*'

c. *hadhaa al-maa-u **kabiir**-un — MSA
this DEF-water-NOM big-NOM
'*This water is big*'

d. *Zhè-xiē shuǐ hěn-**dà** — M. Chinese
this-CL.PL water very-big
'*This water is big*'

Comparison Constructions

Another domain in which construed ontological distinctions integrate with grammar are `COMPARISON CONSTRUCTIONS`, which are discussed at length in *e.g.* Barner and Snedeker (2005); Bale and Barner (2009). In comparison constructions, the interpretative mode of comparison is made on the basis of whether a predicate denotes substances or objects. In cases where the comparison is made with stuff mass terms, the comparison is by `VOLUME`; if the comparison is made with respect to an object mass term, the comparison is by `NUMEROSITY`. The differences between comparison constructions are seen explicitly in (96). Where the stuff mass term WATER is used in (96a), the volume interpretation mode arises, *i.e.*, the amount of water Edward has is judged to be more in volume than the amount of water that Bella has. Where the object mass term FURNITURE is used in (96b), the numerosity interpretation mode arises, *i.e.*, Edward is judged to have more individual chairs, couches, tables *etc.* than Bella. This is the case even if Bella has one large chair with a volume that outnumbers Edward's three small chairs. In other words, the objects denoted by object mass terms are accessible for implicit counting in comparison constructions structures.

(96) a. Edward has more **water** than Bella — English

b. Edward has more **furniture** than Bella — English

The difference in interpretations between object mass and stuff mass in English comparison constructions is mirrored crosslinguistically. In Mandarin Chinese, Welsh, and Hejazi/Modern Standard Arabic collective and general number terms differ from stuff mass terms in their comparison construction interpretations. For a collective or general number term, the interpretative mode is by numerosity (97), and in the case of a stuff mass term, the interpretative mode is by volume (98).

(97) a. M. Chinese

Shīyáng de	**yǐzī**	bǐ	Lù	duō
Shīyáng DE	chair	than	Lu	more

'*Shīyáng has more chairs (in number) than Lu*'

b. Welsh

Mae	gan	Owain	fwy	o	**ellyg**	na	Siwan
be.3	with	Owain	more	of	pear	than	Siwan

'*Owain has more pears (in number) than Siwan*'

c. H. Arabic

Malik	3nd-hu	**tufaaḥ**	'aktr	min	Aknan
Malik	with-the	apple	more	than	Aknan

'*Malik has more apples (in number) than Aknan*'

d. ladii Malik-in **tufaaḥ**-un a'kthr-u min al **tufaaḥ**-i — MSA
with Malik-GEN apple-NOM more-NOM than DEF apple-GEN
alniy ind-a Aknan
(of)-that with Aknan
'*Malik has more apples (in number) than Aknan*'

(98) a. Shīyáng de **shuǐ** bǐ Lù duō — M. Chinese
Shīyáng DE water than Lu more
'*Shīyáng has more water (in volume) than Lu*'

b. Mae gan Owain fwy o **ddŵr** na Siwan — Welsh
be.3 with Owain more of water than Siwan
'*Owain has more water (in volume) than Siwan*'

c. Malik 3nd-hu **muyaa** 'aktr min Aknan — H. Arabic
Malik with-he water more than Aknan
'*Malik has more water (in volume) than Aknan*'

d. ladii Malik-in **ma'a**-un a'kthr-u min al **ma'a**-i — MSA
with Malik-GEN water-NOM more-NOM than DEF water-GEN
alniy inda Aknan
(of)-that with Aknan
'*Malik has more water (in volume) than Aknan*'

Summary of pre-existing perceptual reality

The data shows that on some level pre-existing commitments to perceptual reality, *i.e.* whether a term ultimately denotes objects or stuff, is linguistically relevant and accessible. This is because object mass, collective and general number terms fundamentally differ from stuff mass terms as the objects denoted by object mass, collective and general number terms can be distributed over, count compared, and are relevant for classifier/singulative combinational requirements.

2.2.5 Discussion

The picture painted of object mass, collective and general number terms discussed in this section is summarised in Table 2.12 and Table 2.13. In prose, the data has shown that the collective/singulative systems of Welsh and Hejazi/Modern Standard Arabic at their most basic level share fundamental properties with English object mass systems and Mandarin Chinese general number systems. I submit that object mass, collective and general number systems should therefore be subsumed under one label: `NUMBER NEUTRAL UNCOUNTABLE`, where the primitive features are number neutrality, infelicity when combining with numerals, and a commitment to perceptual reality where objecthood is maintained.

(99) `NUMBER NEUTRAL UNCOUNTABLE`
Terms which in their basic interpretation are number neutral, uncountable, and denote (sums of) pre-theoretical objects. Number neutral uncountable terms includes:

a. `OBJECT MASS TERMS` in number marking languages
e.g. in English

b. `COLLECTIVE TERMS` in singulative languages
e.g. in Welsh and Hejazi/Modern Standard Arabic

c. `GENERAL TERMS` in generalised classifier languages
e.g. in Mandarin Chinese

Following this, the individuating singulative and general classifier are the same primitive function which combines only with a number neutral uncountable term, maintains commitments of objecthood, and derives a (singular) countable term which then obeys the language-specific agreement patterns when combining with numerals.

Finally, the packaging singulative and portioning measures such as QUANTITY are the same primitive function which combines with a stuff mass term, creates individuals in context, and outputs a singular count term with respect to the newly created individuals. As a singular count term, packaging singulatives follow language-specific agreement patterns of singular terms when combining with numerals.

Type	**Refer cumulatively?**	**Combine with numerals?**	**Ontological commitment**
Count Terms			
Singular Count	✗	✔	Objects
Singulatives	✗	✔	Objects
Plural Count	✔	✔	Objects
Bare General Classifier	✗	✔	Objects
Substance Mass Terms			
Subs. Mass	✔	✗	Stuff
Number neutral uncountable terms			
Obj. Mass (English)	✔	✗	Objects
Collectives (Welsh, Arabic)	✔	✗	Objects
General Number (Mandarin)	✔	✗	Objects

Table 2.12: Summary of count, stuff mass and number neutral uncountable terms

	Input	**Output**	**Ontological commitment**
Individuating singulative	Collective	Singular count	Pre-existing
Packaging singulative	Substance Mass	Singular count	Created

Table 2.13: Summary of individuating and packaging singulative features

Points of deviation

Singulative systems, object mass systems and general number systems share the same primitive semantic features, and are so subsumed under the `NUMBER NEUTRAL UNCOUNTABLE` label. There are however some features where singulative systems and canonical classifier systems differ. This short section briefly discusses linear order and syntactic distributions, concluding that such differences are phenomena adjacent to the fundamental number neutral uncountable status of the collective/singulative system.

It will not have escaped the careful reader that Welsh and Hejazi/Modern Standard Arabic singulatives are different from classifiers in Mandarin Chinese in one descriptively obvious way: singulatives are suffixal with a [(num) + N + CL] order and Mandarin Chinese classifiers are independent morphemes with a [(num) + CL + N] order. The differing linear orders are not reason to discount singulative suffixes and classifiers as fundamentally the same. In fact, variation in linear order for classifier structures is typical crosslinguistically. It has been observed in the classier literature(*e.g.* Simpson, 2008; Yi, 2009; Zhang, 2013, *a.o.*) that some languages may have multiple linear orders for classification while other languages only have one canonical order. For example, it has been reported that Korean and Japanese both have the [N + num + CL] order alongside a [(num) + CL + N] order, while for Mandarin Chinese only [(num) + CL + N] is possible, and in Thai only [N + num + CL] is possible. Considering this, it would seem that Welsh and Arabic are languages with two classifier linear orders available: the singulative construction and the pseudo partitive construction.

(100) Singulative construction: [(num) + N + CL]

(bum)	hwyad-en	Welsh
(five)	duck-SING	
'*five ducks*'		

(101) Pseudopartitive construction: [num + CL + N]

bump	o	hwyaid	Welsh
five	of	duck	
'*five ducks*'			

While we may suggest Welsh and Arabic have multiple structures available for the realisation of classifier-like phrases, they nonetheless differ from canonical classifier languages with multiple numeral-classifier surface orders. In Simpson (2008) it is suggested that the default merge order of classifier structures is [num + CL + N], and all other surface orders are derived from this default position by raising. If this is correct, then the default order in singulative languages is the pseudopartitive order (101), and the singulative construction is derived via raising. This being the case, what is the motivation for raising? Simpson suggests that pragmatic motivation licences the choice of specific orders. However, there is no clear semantic or pragmatic interpretative difference between the Welsh or Hejazi/Modern Standard Arabic pseudopartitive and the singulative. It would then seem unlikely that pragmatic raising explains the distributive differences between singulative and pseudopartitive constructions. Instead, the answer may be purely grammatical in nature, where singulative structures allow raising due to case-licensing requirements (as in *e.g.* Longobardi, 1996), though I do not explore this further.

Another point of deviation for singulative markers and canonical classifiers in *e.g.* Mandarin Chinese is syntactic distribution. In Mandarin Chinese, there is an asymmetry between classifier phrases in subject and object position. As shown by Yang (2001), a classifier may be 'stranded' (*i.e.* appear without a numeral) in object position (102a). Yet, in sentence initial contexts (102b), or when the object is separated from the verb (102c), the stranding of a classifier in Mandarin Chinese is strictly banned.

(102) a. Yuehan mǎi-le (yi) ben shu — M. Chinese
John buy-PST one CL book
'*John bought a book*'
From Yang (2001), adapted

b. *(yi) ben shu bu gou — M. Chinese
One CL book not enough
'*One book is not enough*'
From Yang (2001)

c. Yuehan song-le yi gi pengyou *(yi) ben shu — M. Chinese
John gave-PST one CL friend one CL book
'*John gave a friend a book*'
From Yang (2001), adapted

This strict locality condition does not appear on English classifiers nor Welsh and Hejazi/Modern Standard Arabic singulative marked terms, which can appear freely without a numeral in object position as already exemplified by (35)-(37), but also in ditransitive constructions (103), double-object constructions (104)[47], or in subject position.

(103) a. Edward gave **a piece of furniture** to Bella — English

b. Rhoddodd Owain **ellyg-en** i Siwan — Welsh
Give.3.PST Owain pear-SING to Siwan
'*Owain gave a pear to Siwan*'

c. Malik 'a3ṭa **tufaaḥ-ah** li-Aknan — H. Arabic
Malik give.PST.M.3 apple-SING to-Aknan
'*Malik gave an apple to Aknan*'

d. 'a3ṭa Malik-u **tufaaḥ-at**-an li-Aknan — MSA
give.PST.M.3 Malik-NOM apple-SING-ACC to-Aknan
'*Malik gave an apple to Aknan*'

(104) a. Edward gave Bella **a piece of furniture** — English

b. Malik 'a3ṭa Aknan **tufaaḥ-ah** — H. Arabic
Malik give.PST.M.3 Aknan apple-SING
'*Malik gave an apple to Aknan*'

c. 'a3ṭa Malik-u Aknan-in **tufaaḥ-at**-an — MSA
give.3.PST.M.3 Malik-NOM Aknan-gen apple-SING-ACC
'*Malik gave Aknan an apple*'

[47] There is no double object construction parallel to English in Welsh *Owain gave Siwan a fish*: **Rhoddodd Dafydd Siwan bysgodyn.*

(105) a. Bwytodd **hwyad-en** fwydyn — Welsh
ate.PST.3 duck-SING worm
'*A duck at a worm*'

b. **baṭ-ah** 'akalat duud-ah — H. Arabic
duck-SING eat.PST.F.3 worm-SING
'*A duck ate a worm*'

c. 'akalat **baṭ-at**-un duud-at-in — MSA
eat.PST.F.3 duck-SING-NOM worm-SING-ACC
'*A duck at a worm*'

Again, there is no reason to take these distributional differences as evidence that classifiers and singulatives are not akin. As discussed by Cheng and Sybesma (1999), Cantonese, another canonical classifier language, experiences sentence initial [noun + CL] constructions. However, the Cantonese [noun + CL] string is restricted to definiteness, a feature not seen in Mandarin Chinese, Welsh, Hejazi/Modern Standard Arabic nor English.

(106) Zek gau gamjat dakbit tengwaa — Cantonese
CL dog today special obedient
'*The dog is specially obedient today*'
From Cheng and Sybesma (1999, p.511)

With this brief discussion, I take general classifiers and singulatives as semantically comparable in their basic function, *i.e.* they create a semantic singularity which maintains pre-existing perceptual commitments of the number neutral uncountable term it combines with and allows further combination with numerals. I will assume that distributional asymmetries and (non)-obligatory definiteness are adjacent issues not inherently related to classifier systems.

2.3 Bare Plurals

Singulative marked terms are semantic singulars and descriptively differ from singular count terms only in their morphological markedness status. It is reasonable to expect singulatives to behave as singulars in all environments. We have so far seen that this is the case when combining singulative marked terms with numerals: language specific agreement patterns which apply to singulars also apply to singulatives.

This section explores a case where singulars and singulatives seemingly diverge in behaviour: bare pluralisation (pluralisation without numerals). At first glance, the behaviour of the bare plural of the singulative does not mirror the behaviour of the bare plural of the singular in either Welsh nor Hejazi/Modern Standard Arabic. In this section, I show that in Arabic this is an illusion due to homophony of morphological form with paucal semantics, but in Welsh there is truly an asymmetry, as there is no (overt) plural of the singulative.

Before introducing the properties of the bare plural of the singulative, let us briefly explore what is known about the semantics of the bare plural of singular terms. Enquiry into the properties of plural marking is no trivial endeavour, and has sparked lively scholarly debate. Common wisdom indicates that the semantic difference between singular and plural-marked terms is that singulars denote *one* and plural marked terms denote *more than one*. Indeed, this is such a prevailing and intuitively obvious conclusion that it is taken as the definition of `PLURAL` even in dictionaries dedicated to linguistic phenomena.

> **Plural**
> Definition: Feature of forms used in referring to more than one, or more than some small number of individuals. Often a term in an inflectional category of number: *e.g.* plural rooms ('room-pl') is distinguished from the singular room, by -s as a plural ending. But the meaning may also be that of an independent particle, or of a derivational affix.
> From *The Concise Oxford Dictionary of Linguistics* (Matthews, 2014)

This simple definition of plural obviously holds in upwards entailing contexts in English, Welsh and Hejazi/Modern Standard Arabic, as shown in the repeated examples (38)-(41).[48] Explicitly, these statements are false in the case where a single girl was seen in the garden. As such, the meaning of the plural marker in these contexts is sum-based, and excludes reference to singular entities.

(38) Edward saw **girls** in the garden — English

(39) Gwelodd Owain **ferch-ed** yn yr ardd — Welsh
see.PST.3 Owain girl-PL in DEF garden
'*Owain saw girls in the garden*'

(40) Malik shaaf **ban-aat** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 girl-PL in DEF-garden
'*Malik saw girls in the garden*'

[48] Mandarin Chinese is excluded from this discussion as the status of plural markers is controversial and it is not immediately clear which plural markers parallel English, Welsh and Hejazi/Modern Standard Arabic plural markers. For example, Mandarin Chinese has the plural suffix -MEN, which has been claimed to have meanings outside of simple sum formation, such as associativity (Park, 2008; Cheng and Sybesma, 1999; Kim, 2008; Li, 1999; Kang, 1994; Iljic, 1994). Not only is -MEN linked to associativity, it is restricted to definite human (or human-like) terms (Zhang, 2013), and the first person singular pronoun (Harbour, 2014; Zhang, 2013). Another candidate for Mandarin Chinese plural markers is XIE which induces a strictly exclusive interpretation (Iljic, 1994; Li, 1999; Zhang, 2013; Wu, 2019), or reduplication of the classifier as in (iii), which also has a strict exclusive interpretation (Zhang, 2013). I briefly return to reduplicated classifiers in §6.3.2.4.

(i) Shīyáng zài huāyuán lǐ kàndào-le **nǚhái-men** — M. Chinese
Shīyáng at garden in see-PST girl-MEN
'*Shīyáng saw a group of girls in the garden*'

(ii) zuótiān wo kanjian le yi **xie** xuesheng. Wo qing tamen chi fan le — M. Chinese
yesterday 1 see PST one xie student 1.SING invite 3.PL eat rice PST
'*Yesterday I saw some students. I treated them to a meal*'
From Wu (2019, p.11), adapted

(iii) He-lǐ piao-zhe (yi) pian-pian shuye — M. Chinese
river-in float-DUR one CL-RED leaf
From Zhang (2013, p.3)

(41) ra'aa Malik-u **fatay-aat**-in fi al-hadiqat-i MSA
see.PST.M.3 Malik-NOM girl-PL-GEN in DEF-garden-GEN
'*Malik saw girls in the garden*'

Yet, if plural marking indicates a meaning of *two or more*, then it is difficult to explain why singular reference is not always excluded for plural marked terms, such as in downward entailing environments. When placed in a standard downwards-entailing context, the English plural is naturally interpreted as inclusive of the singular (see Farkas and de Swart, 2010; Ivlieva, 2013; Martí, 2017; Mayr, 2015; Sauerland et al., 2005; Spector, 2007; Zweig, 2009, *a.o.*). In (107), it is not the case that Edward did not see *two or more* girls in the garden. Rather, it is the case that Edward did not see *one or more than one* girl in the garden. Similar observations are seen for the Welsh and Hejazi/Modern Standard Arabic plurals. Going forward, I refer to plurals which *exclude* reference to singular entities, as in (38)-(41) as `EXCLUSIVE PLURALS`, and plurals which *include* reference to singular entities, as in (107)-(110) as `INCLUSIVE PLURALS`.

(107) Edward didn't see (any) **girls** in the garden English

(108) Ni welodd Owain **ferch-ed** yn yr ardd Welsh
NEG see.PST.3 Owain girl-PL in DEF garden
'*Owain didn't see girls in the garden*' (inclusive)

(109) maa shaaf Malik **ban-aat** fi al-hadiqa H. Arabic
NEG see.PST.M.3 Malik girl-PL in DEF-garden
'*Malik didn't see girls in the garden*' (inclusive)

(110) lam yara Malik-u **fatay-aat**-in fi al-hadiqat-i MSA
NEG see.JUSS.M.3 Malik-NOM girl-PL-GEN in DEF-garden-GEN
'*Malik didn't see girls in the garden*' (inclusive)

Plural marked terms are interpreted inclusively not only in simple negated environments, but also in other downward entailing environments, such as conditionals. In examples (111)-(114) even though the plural form is used, if the instructions are followed correctly, then a report should be made in the instance a singular girl was seen in the garden.

(111) If you see (any) **girls** in the garden, English
please notify a member of staff

(112) Os gwelwch **ferch-ed** yn yr ardd, Welsh
If see.imp.2.pl girl-pl in def garden
rhowch wybod i aelod o staff.
give.imp.2.pl knowledge to member of staff
'*If you see any girls in the garden, please tell a member of staff*'

(113) mumkin 'idhaa shuft **ban-aat** fi al-hadiqah, H. Arabic
please if see girl-pl in def-garden
taquul li-wahid min al 3amliin
say to-one of def staff
'*If you see any girls in the garden, please tell a member of staff*'

(114) 'idhaa ra'ayta **fatay-aat-in** fi al-hadiqat-i, MSA
if see.pst.m.2 girl-pl-gen in def-garden-gen
halaa tafaḍalta bi'ikhbaar a'hadi al-3amliin
do prefer.2 with-news one-of def-staff
'*If you see any girls in the garden, please tell a member of staff*'

So too are plural marked terms interpreted as inclusive in question environments. In (115)-(118), the questions are asking if *any* amount of girls are in the flowerbed, including *one*. As such, the questions may be felicitously answered in the case that a single girl is in the flower bed.

(115) Q: Are there girls in the garden? English
A: Yes, there's **one** sitting in the flower bed

(116) Q: Oes yna **ferch-ed** yn yr ardd? — Welsh
be.3 there girl-PL in DEF garden?
'*Are there girls in the garden?*'

A: Oes, mae yna **ferch** yn eistedd yn y gwely blod-au.
Yes, be.3 there girl in sitting in DEF bed flower-PL
'*Yes, there a girl sitting in the flower bed*'

(117) Q: huwa fi **ban-aat** fi al-hadiqa? — H. Arabic
there in girl-PL in DEF-garden?
'*Are there girls in the garden?*'

A: aywa, fi **bint** wahida qaa'da 'inda hawḍ al-a'zhar-i
yes, in girl one-F sit in basin DEF-flower-GEN
'*Yes, there a girl sitting in the flower bed*'

(118) Q: hal hunaak-a ayy **fatay-aat**-in fi al-hadiqat-i? — MSA
Q there-F any girl-PL-GEN in DEF-garden-GEN
'*Are there girls in the garden?*'

A: na'am, hunaak-a **fatay**-un waahid-un tajlisu 'inda hawd al-a'zhar-i
yes, there-F girl-NOM one-NOM sit.F.3 in basin DEF-flower-GEN
'*Yes, there a girl sitting in the flower bed*'

Considering that singulative marking derives a singular count term, then it is reasonable to expect the plural of the singulative to behave as a plural of the singular in relevant contexts. As it turns out, the picture is not so simple. The Arabic plural of the singulative is indeed attested, but is typically reported as having a paucal interpretation (Mathieu, 2014; Dali and Mathieu, 2020; Jaradat and Jarrah, 2022), a meaning not inherently associated with the plural of the singular. In Welsh, the plural of the singulative is completely absent. If singulatives are singulars subject to functions upon singulars, as I contest they are, then both these results are unexpected. I turn to both these issues independently.[49]

[49] Singulatives in other languages are said to also pluralise, including Ojibwe (Mathieu, 2012a,b), Maltese (Acquaviva, 2015; Dali and Mathieu, 2021a), Breton (Acuaviva, 2008; Acquaviva, 2015; Dali and Mathieu, 2021a; Grimm, 2012b,a; Mathieu, 2012a,b), Russian (Wierzbicka, 1988). I have little to say about these languages, except Breton, which I return to in chapter 6.

2.3.1 The Arabic plural of the singulative

The status of the Arabic plural of the singulative has already received some attention in the literature. While it is known that Arabic individuating singulative can pluralise, its status as a plural parallel to plurals of singular terms is contested. Borer and Ouwayda (2010, 2021) report that the bare Arabic feminine sound plural of the singulative is ungrammatical, and in context of numerals it is merely agreement. This is not the commonly accepted view. For Dali and Mathieu (2021a,b); Mathieu and Dali (2021); Fassi Fehri (2016); Mathieu (2012a, 2013, 2014); Ojeda (1992); Zabbal (2002) the Arabic feminine sound plural of the singulative is not simply agreement, as it may appear bare (without numerals).[50] The general claim is that though felicitous, the semantic status of the bare feminine sound plural of the singulative does not parallel the feminine sound plural of the singular in all contexts. Cowell (1964); Dali and Mathieu (2021a,b); Jaradat (2023a,b) and Mathieu and Dali (2021) have already observed that bare plurals of singulatives in upwards entailing contexts are indeed interpreted exclusively, but with a paucal interpretation in a number of Arabic dialects (*e.g.* Syrian, Jordanian). The same observation holds for both Hejazi Arabic and Modern Standard Arabic, as shown in in (119) and (120), where the interpretation is that a small amount of ducks/trees, typically less than ten, were seen. Note, I gloss the feminine sound plural of the singulative as paucal (PAU) here for clarity.

(119) Malik shaf **baṭ-a-at/shajar-a-at** H. Arabic
Malik see.PST.M.3 duck-SING-PAU/tree-SING-PAU
fi al-hadiqa
in DEF-garden
'*Malik saw a small number of ducks/trees in the garden*'

[50] Note, these accounts only discuss the feminine sound plural of the singulative, with no discussion of the *broken* plural of the singulative.

(120) ra'aa Malik-u **baṭ-a-at-in/shajar-a-at-in** MSA
see.PST.M.3 Malik-NOM duck-SING-PAU-ACC/tree-SING-PAU-ACC
fi al-hadiqat-i
in DEF-garden-GEN
'*Malik saw a small number of ducks/trees in the garden*'

The status of the feminine sound plural of the singulative in downward entailing contexts and questions has only been investigated by Éric Mathieu and Maryam Dali, though there is inconsistency with the reporting of interpretations. Mathieu (2014) reports that the bare feminine sound plural of the singulative is ungrammatical in downward entailing and question environments while Dali and Mathieu (2021b); Mathieu and Dali (2021) reports the feminine sound plural of the singulative in downward entailing contexts as grammatical only if interpreted exclusively and paucally. In Hejazi/Modern Standard Arabic as it is spoken by the consultants of this thesis, the status of the feminine sound plural of the singulative in simple negative (121), (122), conditional (123), (124) and question environments (125), (126) is marginal at best, preferably with a paucal interpretation. Though the available evidence suggests the acceptance of the bare feminine sound plural of the singulative is not clear-cut (and perhaps careful experimentation is required to reveal its true nature), I conclude that in downward entailing environments the feminine sound plural of the singulative is not free in the same manner as the plural of the singular.

(121) #maa shaaf Malik **baṭ-a-at** fi al-hadiqa H. Arabic
NEG saw.PST.M.3 Malik duck-SING.PAU in DEF-garden
Intended: '*Malik didn't see ducks in the garden*'

(122) #lam yara Malik-u **baṭ-a-at**-in MSA
NEG see.JUSS.M.3 Malik-NOM duck-SING.PAU-GEN
fi al-hadiqat-i
in DEF-garden-GEN
Intended: '*Malik didn't see ducks in the garden*'

(123) #mumkin 'Idhaa shuft ayy **baṭ-a-at** fi al-hadiqah, H. Arabic
please if see any duck-SING-PAU in DEF-garden

taquul li-wahid min al 3amliin
say to-one of DEF staff
Intended: '*If you see ducks in the garden, please tell a member of staff*'

(124) #'idhaa ra'ayta **baṭ-a-at-in** fi al-hadiqat-i, MSA
if see.PST.M.2 duck-SING-PL-GEN in DEF-garden-GEN

halaa tafaḍalta bi'ikhbaar a'hadi al-3amliin
do prefer.2 with-news one-of DEF-staff
Intended: '*If you see ducks in the garden, please tell a member of staff*'

(125) #huwa fi **baṭ-a-at** fi al-hadiqa? H. Arabic
there in duck-SING-PL in DEF-garden?
'*Are there ducks in the garden?*'

(126) #hal hunaaka ayy **baṭ-a-at-in** fi al-hadiqat-i? MSA
Q there-F any duck-SING-PL-GEN in DEF-garden-GEN
'*Are there girls in the garden?*'

The fact that the Arabic feminine sound plural of the singulative is restricted has lead to claims that it is inherently paucal and exclusive in a way that the plural of the singular is not (Dali and Mathieu, 2021a,b; Mathieu and Dali, 2021; Jaradat, 2023a; Jaradat and Jarrah, 2022). Discussions surrounding the availability of paucal readings in (the dialects of) Arabic spans beyond what is seen for the behaviour of the feminine sound plural of the singulative. Inquiry into the intricacies of paucal number in Arabic goes back to at least Wright (1896, p.234) who reports that the interpretation of paucal number is dependent upon the existence of competing plural forms. According to Wright, certain (broken) plural forms are generally associated with paucal interpretations only if there are other non-paucal plural forms with which it contrasts. In the case that there is only one plural form of a singular, then the paucal interpretation does not arise.

While Wright's grammar is now dated, similar observations have been reported more recently by Cowell (1964, p.369), Jaradat (2023a) and Dali and Mathieu (2021b, §4.3). For these authors, across dialects of Arabic, it is in fact the feminine sound plural which assumes the paucal interpretation, but *only* when there is broken plural form which it contrasts against.[51] The example Jaradat (2023a) gives to exemplify this pattern is singular count KHATIYAH (sin.SG = *'sin'*), with feminine sound plural KHATIYA-AT (sin.SG-PL = *'sins'*) which is interpreted as paucal, and the broken plural KHATAAYA (sin.SING.BRPL = *'sins'*) which is not interpreted as paucal.[52]

> feminine sound plural of a count nominal can be paucal if a broken plural template can also pluralise that nominal.
> (Jaradat, 2023a, p.7)

The fact that the feminine sound plural is not interpreted as paucal when there is no alternative plural form is shown more explicitly in (127). In most dialects of Arabic, there is only one plural for singular count DARAAJA (bike.SG = *'bike'*), which is the sound plural DARAAJA-AT (bike.SG-PL = *'bikes'*). In the examples in (127) and (128) there is no paucal interpretation of the feminine sound plural, as there is no alternative (broken) plural form for it to contrast against.

(127) Malik shaaf **daraaja-at** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 bike-PL in DEF-garden
'*Malik saw bikes in the garden*'

(128) ra'aa Malik-u **daraaja-at-in** fi al-hadiqat-i — MSA
see.PST.M.3 Malik--NOM bike-PL-GEN in DEF-garden-GEN
'*Malik saw bikes in the garden*'

[51] Interestingly, Dali and Mathieu (2021b, p.73) discuss the feminine sound plural/broken plural contrastive interpretations of the singular/plural domain (feminine sound plural is paucal only if contrastive), but they do not extend the discussion to the collective/singulative domain.

[52] Transliteration adapted from Jaradat (2023a).

I submit that the paucal interpretations of the Hejazi/Modern Standard Arabic feminine sound plural of the singulative are reducible to general patterns of Hejazi/Modern Standard Arabic paucal marking and interpretations. The thesis I put forward is that the feminine sound plural of the singulative is the canonical realisation of the PAUCAL OF THE SINGULATIVE. In these cases, feminine sound plural realisation is expected, as there is a contrast against another plural form: the collective. That it is a contrast with a collective is key; (119) has the paucal interpretation even though BAṬ-AH has no broken plural to contrast against.

(119) a. Malik shaaf **baṭ-a-at/shajar-a-at** H. Arabic
Malik see.PST.M.3 duck-SING.PAU/tree-SING.PAU
fi al-hadiqa
in DEF-garden-GEN
Malik saw a small number of ducks/trees in the garden

In a sense, Dali and Mathieu (2021a,b); Mathieu and Dali (2021); Jaradat (2023a) and Jaradat and Jarrah (2022) are correct: there is a semantics associated with the feminine sound plural of the singulative that is in some way inherently paucal. The question is, then: is there a plural of the singulative that is not inherently paucal, *i.e.* parallel to the plural of the singular, including appropriate inclusive and exclusive interpretations when bare? I submit this is one of the interpretations of the broken plural of the singulative, where available.

In the literature that concerns itself with the Arabic collective/singulative systems, there is little discussion or acknowledgement, to my knowledge, outside the works of Fassi Fehri (2004, 2012) of the broken plural of the singulative.[53] As we have already seen in §2.2.2, the broken plural of the singulative may be used in numeral contexts below ten. When bare, the broken plural of the singulative facts are obfuscated by the fact that it is formally identical to a plural of the collective as in (129) and (130) which have multiple interpretations. For the broken plural of the

[53]Borer and Ouwayda (2021) and Hnout et al. (2021) acknowledge but do not discuss the broken plural of the singulative.

singulative reading (129) and (130) are interpreted exclusively as a plurality of trees with no paucal interpretation.[54]

(129) Malik shaaf **'ashjaar** fi al-hadiqa — H. Arabic
see.pst.m.3 Malik tree.sing.pl in def-garden
Plural reading: '*Malik saw trees in the garden*'
Abundance reading: '*Malik saw many trees in the garden*'

(130) ra'aa Malik-u **'ashjaar-in** fi al-hadiqat-i — MSA
see.pst.m.3 Malik-nom tree.sing.pl-acc in def-garden-gen
Plural reading: '*Malik saw trees in the garden*'
Kind reading: '*Malik saw kinds of trees in the garden*'
Abundance reading: '*Malik saw many trees in the garden*'

Placing the broken plural of the singulative in downward entailing contexts (131)-(134), conditionals (133)-(134) and questions (135)-(136) is not only grammatical, but also not restricted to paucal interpretation. Rather, they are interpreted as an inclusive plural, in line with plural of singular interpretations in these contexts.

(131) maa shaaf Malik **'ashjaar** fi al-hadiqa — H. Arabic
neg see.pst.m.3 Malik tree.sing.pl in def-garden
'*Malik didn't see trees in the garden*'

(132) lam yara Malik-u **'ashjaar-in** fi al-hadiqat-i — MSA
neg see.juss.m.3 Malik-nom tree.sing.pl-gen in def-garden-gen
'*Malik didn't see ducks in the garden*'

[54]Fassi Fehri (2004, p.62-63) notes similar observations for the broken plural of the singulative.

(133) mumkin ‘Idhaa shuft **’ashjaar** fi al-hadiqah, H. Arabic
please if see tree.SING.PL in DEF-garden,

taquul li-waahid min al-eaamliin
say to-one of DEF-staff
‘*If you see ducks in the garden, please tell a member of staff*’

(134) ‘idhaa ra’ayta **’ashjaar-in** fi al-hadiqat-i, MSA
if see.PST.M.2 tree.SING.PL-GEN in DEF-garden-GEN

halaa tafaDalta bi’iikhbar a’hadi al-eaamliin
do prefer.2 with-news one-of DEF-staff
‘*If you see trees in the garden, please tell a member of staff*’

(135) huwa fi **’ashjaar** fi al-hadiqa? H. Arabic
there in tree.SING.PL in DEF-garden?
‘*Are there trees in the garden?*’

(136) hal hunaak-a **’ashjaar-in** fi al-hadiqat-i? MSA
Q there-F tree.SING.PL-GEN in DEF-garden-GEN
‘*Are there trees in the garden?*’

These broken plural of the singulative facts offer a more complete picture of the interaction between plural markers and the collective/singulative system of Hejazi/Modern Standard Arabic. It is not the case that the plural of the singulative must be exclusive and paucal. Rather, as the singulative marked term is a semantic singular, it may be pluralised using the standard plural forming strategies in Hejazi/Modern Standard Arabic. Following standard morphological patterning, standard pluralisation is typically realised as a broken plural (see §2.1.1). When singulatives combine with a paucal, it is necessarily realised as a feminine sound plural because of the contrasting collective form.

2.3.2 The absence of the Welsh plural of the singulative

Welsh has no plural of the individuating singulative, bare or otherwise. Ungrammaticality of the Welsh plural of the singulative may be expected in numeral contexts, as in (137), as Welsh numerals always combine with the formally singular marked term. This does not explain the absence of the bare plural of the singulative: the example in (138) is outright ungrammatical. As singulatives are singular terms, this is unexpected.

(137) *Gwelodd Owain bum **hwyad-enn-au** yn yr ardd — Welsh
see.PST.3 Owain five duck-SING-PL in DEF garden
Intended: '*Owain saw five ducks in the garden*'

(138) *Gwelodd Owain **hwyad-enn-au** yn yr ardd — Welsh
see.PST.3 Owain duck-SING-PL in DEF garden
Intended: '*Owain saw ducks in the garden*'

The absence of the bare Welsh plural of the singulative cannot be due to any rule that uniformly disallows pluralisation of classified or measure phrases. The application of the plural marker to the measure phrases in (139) is legal, allowing the interpretation of sums of bottles which contain water. Similar behaviour is seen for the packaging singulative in (140), where the interpretation is of multiple, non-overlapping, pieces of cheese.

(139) Prynodd Owain **potel-i o ddwr** — Welsh
buy.PST.3 Owain bottle-PL of water
'*Owain bought a bottle of Water*'

(140) Bwytodd Owain **gos-ynn-au** ddoe — Welsh
eat.PST.3 Owain cheese-PSING-PL yesterday
'*Owain ate pieces of cheese yesterday*'

This absence of the bare plural of the singulative in Welsh is not only unexpected, but it is the single point of deviation for the Welsh and Arabic collective/singulative systems. As far as I am aware, the fact that the Welsh individuating singulative does not pluralise is only explicitly noted in the literature by Nurmio (2017, 2023). It is common to find literature that either incorrectly assumes or uses inaccurate data to claim that the Welsh individuating singulative pluralises.[55] For example, Dali and Mathieu (2021b) use the COSYN and GLÖYN paradigms in Table 2.14 as evidence that the singulative in Welsh pluralises.[56] However, COS-YN (caws.SMASS-PSING = *'piece of cheese'*) is an instance of a packaging singulative, and is semantically distinct from an individuating singulative. The same can also be said of GLÖ-YN, where GLO (*'coal'*) refers to the mass substance, and GLÖ-YN is a partition of that substance. We therefore have an asymmetry: the Welsh individuating singulative does not pluralise but the packaging singulative may do so.

Mass	**Singulative**	**PoTS**	**Gloss**
caws	cos-yn	cos-ynn-au	cheese → piece of cheese → pieces of cheese
glo	glö-yn	gloo-ynn-au	coal → piece of coal → pieces of coal

Table 2.14: Welsh plural of the packaging singulative marking

Before moving on, there is one example of a Welsh plural of the singulative in the literature, used by Grimm (2012b), which seems to throw doubt onto the claim that individuating singulatives do not pluralise: the GRONYNNAU paradigm (141).

(141) grawn → gronyn → gronynnau Welsh
grain → a grain → grains

[55] In defence of Dali and Mathieu (2021b), their focus is Arabic and they use Welsh data as supplementary evidence for their primary interests of enquiry. For Grimm (2012b), their focus is Welsh, but their examples come from Stolz (2001), who lists both GRONYNNAU and COSYNNAU as a plural of the singulative. No other plural of the singulative are listed in Stolz's paper. Other often cited sources in the literature for the apparent Welsh plural of the singulative are Cuzzolin (1998), who also offers the COSYN paradigm.

[56] *c.f.* Heinz (2009) who cites the plural of COSYN as COSYNNOD

Unlike COSYN and GLÖYN, this example is not clearly a packaging singulative. Instead, it looks like an individuating singulative, where the individual, namely a seed, is a bounded object in relation to a collective group of like-aggregates, namely seeds. Therefore, it would seem that GRONYNNAU is the plural of the individuating singulative, a direct contradiction to the claim that both Nurmio (2017, 2023) and I argue is not attested in Welsh. To this, I submit that the GRONYN paradigm is evidence for a Welsh plural of the singulative. Instead, I suggest that GRONYNNAU is not a plural of individuating singulative GRONYN, but rather a plural of GRONYN as a *re-lexicalised singular*. As such, there is a homophonic pair with two distinct meanings, leading to two paradigms: a collective/singulative pair as in (142) and a singular/plural pair as in (143).

(142) grawn → gronyn — Welsh
grain → a grain

(143) gronyn → gronynnau — Welsh
particle → particles

I argue for the dual pair as there is no evidence that GRONYNNAU and GRAWN are part of the same inflectional paradigm in modern Welsh in the way that, for example, SHAJAAR and ASHJAAR are clearly in the same paradigm in Arabic.[57] The reasoning for this is that both collective SHAJAAR (tree.COL = *'tree(s)'*) and broken plural of the singulative 'ASHJAAR (tree.COL.SING.PL = *'trees'*) in Hejazi/Modern Standard Arabic ultimately refer to like-individuals *trees*, the difference between these terms is grammatical. On the other hand, the difference between GRAWN and GRONYNNAU is not only grammatical, but also conceptual as the refer to fundamentally different types of individuals. To see this, consider that the following examples where GRAWN refers to *cereals/seeds*, and GRONYNNAU is closer in meaning to *particles* or *granules*. As there is no such individual as an 'atomic seed', the use of GRAWN in (145) is ruled out, and only GRONNYNAU as the plural of re-lexicalised GRONYN is acceptable.

[57] This of course offers no claim to any historical relatedness nor diachronic development.

(144) mae atomau wedi eu gwneud o **ronynnau is-atomig** Welsh
'*Atoms are made of sub-atomic particles*'

(145) # mae atomau wedi eu gwneud o **rawn is-atomig** Welsh
Intended: '*Atoms are made of sub-atomic particles*'

Similarly, in (146) and (146) GRONYNNAU ensures grammaticality, while GRAWN does not. I submit this difference is due to the semantic infelicitousness of GRAWN: there are no such individuals as coffee cereals.

(146) Tywalltodd Owain y **gronynnau coffi** lawr y sinc Welsh
'*Owain poured coffee granules down the sink*'

(147) #Tywalltodd Owain y **grawn coffi** lawr y sinc Welsh
Intended: '*Owain poured coffee granules down the sink*'

The hypothesis that GRONYNNAU is not derived from GRAWN in modern Welsh is supported by corpus data. In *The Proceedings of National Assembly for Wales Corpus*[58] the only instances of GRONYNNAU refer to *particles* rather than seeds or cereals. The two examples and translations in (148) and (149) are lifted directly from the corpus, but the emphasis is my own.

(148) Dylai fod yn fonitro parhaus sy'n caniatáu i'r cyhoedd yn yr ardal leol gael rhan yn y gwaith o fonitro'r **gronynnau** a'r llygredd yn yr atmosffer. Welsh
'*It should be a continual monitoring that allows the public in the local area to be engaged in monitoring the* ***particulate*** *and pollution in the atmosphere.*'

[58]Corpora accessible via Bangor university. Available at:
https://corpws.cymru/proceedings-of-the-national-assembly-for-wales/?lang=en#cy_gronynnau
Link accessed 29th January 2024.

(149) Gobeithiaf y byddwn, gyda'n gilydd, yn gallu rhoi sicrwydd i gymunedau yng Nghastell-nedd a Phort Talbot y byddwn yn ceisio cydweithio i sicrhau ein bod yn cadw lefelau'r **gronynnau** mor isel ag sy'n bosibl. Welsh

'*I hope that, together, we will be able to reassure the communities in Neath and Port Talbot that we will look to work together on ensuring that we keep the **particulate** levels as low as possible.*'

I have so far claimed that Welsh packaging singulatives may pluralise, but individuating singulatives do not, and any exceptions are either plurals of packaging singulatives or cases that involve re-lexicalisation. There is however one wrinkle: artificial collection singulatives are marginally more acceptable in the plural if the individuality of objects is highlighted/contrasted over the collection.

(150) #Prynodd Owain **ddodrefn-yn-au** ddoe Welsh
buy.PST.3 Owain furniture-SING-PL yesterday
'*Owain bought individual pieces of furniture yesterday*'

(151) 'Er y byddai **dodrefnynnau** a drinir yn unigol yn dod yn is na'r trothwy cyfalafu, at ei gilydd maent yn asedion arwyddocaol ac felly cronwyd nhw at ei gilydd am y tro cyntaf yng nghyfrifon 1999/2000.' Welsh

'*Although **items of furniture**, if treated singly, would fall below the capitalisation threshold, collectively they represent a significant asset and accordingly they were pooled for the first time in the 1999/2000 accounts. The existing stock of furniture was estimated to be half way through its life as at 31 March 2000*' [59]

[59] Example and translation from the Senedd (Welsh government). Available at https://senedd.wales/media/hqsfj5at/bus-guide-3c5a6a2d000774d10000259000000000-english.pdf accessed 12th January 2024.

I take from this that any plurals of the singulative in Welsh are idiosyncratic, extremely marginal, and may be restricted to specific terms within the artificial collection notional category. At the very least, the plural of the singulative does not apply to singulative terms which refer to natural kind notions. Evidence to the contrary in the literature is misleading at best, as often 'plurals of the singulative' are actually cases of plurals of packaging singulatives (COSYNNAU) or plurals of re-lexicalised singulars (GRONYNNAU).

2.4 Summary and next steps

This chapter has been dedicated to a descriptive overview of the characteristics of the Welsh and Hejazi/Modern Standard Arabic number systems, with a focus on a direct comparison of their collective/singulative systems to the object mass system in English and the general number system in Mandarin Chinese. The data shows that collective/singulative, object mass, and general number systems are reducible to the same semantic primitives. Object mass, collective and general number terms are number neutral uncountable terms. The picture painted of singulative marking in general equates singulatives with general classifiers.

Singulatives have been shown to be a macro-category, with at least two subcategories of true singulative functions (as opposed to faux singulatives): packaging singulatives apply to substance mass terms and create a measure/partition in context that is grammatically singular count, and individuating singulatives apply to collectives to name a singular individuals for which we have ontological commitments. By this measure, the packaging singulative is a quantity phrase a la English QUANTITY OF or PORTION OF, and the individuating singulative is a generalised classifier a la Mandarin Chinese GÈ.

As singulative marked terms are singular count terms, they are expected to behave as singular count terms in relevant contexts. With respect to their agreement patterns when combining with numerals, this is borne out. For Arabic, this requires plural agreement with numerals below ten, singular agreement otherwise. For Welsh, this requires singular agreement. With respect to appearing bare, the behaviours of the plural of the singulative in Arabic are as expected, where paucals and plurals of the singulative follow expected agreement patterns. The absence of the Welsh plural of the singulative is however deviant and as yet unexplained.

With this overview in place, it is obvious that any analysis of the Welsh and Hejazi/Modern Standard Arabic collective/singulative systems should primarily be in the context of wider discussions of uncountability and the count/mass distinction a la Grimm (2012b); Mathieu (2012a); Dali and Mathieu (2020), rather than in the spirit of plural allomorphy/allosemy a la Kouneli (2019) and Erschler (2022). The questions still remain: what is the architecture of this system? Should it be lexicalist or constructionist in nature? How can it explain the absence of the Welsh plural of the singulative?

Chapter 3

Parts, Wholes, Objects, Substances, and Distance

The primary aim of this chapter is to ground how we may think of ontological distinctions such as integrated and non-integrated wholes and the substances/object distinction from a metaphysical perspective. The concepts discussed in this chapter will form the foundation of construed ontological modelling (*i.e.* how we perceive the structure of the world) and linguistic ontological modelling (representations of individuals in natural language).

In §3.1, I set the scene of enquiry, suggesting that ontological distinctions concern at the very least parts of wholes, how those parts are connected, and the constituent matter that makes up those wholes.

In §3.2, I discuss the basics of partially ordered sets, Hasse diagrams, lattices and Boolean algebras, as these structures are typically used in formal approaches to mereology and mereotopology. The discussion in this section is algebraic in nature, and is not discussed in light of part-whole relationships.

In §3.3, I discuss the application of partially ordered sets in `MEREOLOGY`, the theory of parthood. The purpose of this section is to introduce a ground mereology that will serve as the basis for discussion of both parts and wholes at a conceptual level, but also theories of nominal number that are based on mereological concepts. The discussion in this section relies on the formalisation and discussion found in Champollion and Krifka (2016); Partee et al. (1990).

In §3.4, I discuss how mereology in and of itself is not sufficient to describe what it means to be an integrated whole, and so introduce mereotopology. The distinction between arbitrary wholes and integrated wholes, with an integrated whole being described as `MAXIMALLY STRONGLY SELF CONNECTED`.

In §3.5 I outline mereotopological notions that can be used to describe the intuitive difference between substances and non-substances. The discussion in this section is original, and grounds my assumptions going forward on the very nature of the substance/object distinction.

In §3.6 I outline how mereotopological notions cannot capture issues of metric distance. I introduce distance functions which will be used to describe measures of space between two individuals.

Much of the formalism presented in this chapter is built on standard formalisms of algebraic concepts and mereo(topo)logy, and as such the presentation of these concepts is also fairly standard. I assume no background knowledge of the reader, and so primitives are presented in their entirety. For advanced readers with familiarisation in these topics, some sections may easily be skipped.

3.1 Setting the scene

Thinking about the metaphysical structure of individuals is to think about their position in space, their parts, the composition of these parts, and how all these parts are connected. To think of a flower as an integrated whole, one must accept that the flower is separate from the non-flower things that may be touching it. The little caterpillar may be sitting on the flower's leaves, and the caterpillar and the flower may be vanishingly close with little perceptual space between them, but the caterpillar and the flower remain separate wholes. Similarly, to think of my cat Klaus as an integrated whole, we must consider Klaus' paws as part of Klaus, but not the mittens he is wearing. Though Klaus is an integrated whole qua the CAT property, at the same time we may consider Klaus' paw as a whole in its own right qua the PAW property, even though the paw is part of Klaus. Expressing these notions in a meaningful way requires formalising what it means to be an (integrated) whole but also what it means to be *part-of* a whole.

The study of parthood is broadly called `MEREOLOGY`, which has a long history in the philosophical tradition. Early discussions of mereological concepts date back to Aristotle's *Metaphysics* book Δ, but it was in the early 20th century that formalism of mereological concepts emerged, starting with Leśniewski (1916) in Polish, and being further developed by Leonard and Goodman (1940), Goodman and Quine (1947) and Goodman (1951) in English. As a field of enquiry, mereology includes not only physical *part-of* relations (such as *Klaus' paw* being part of *Klaus*), but also plural membership (*John* being part of *John and Bill*), grouphood membership (*Liz Truss* being a member of *the House of Commons*[1]), events and subevents (*my writing this thesis* as part of *my completing a PhD*).

Strictly, mereology in and of itself is a theory of parts and wholes, and does not distinguish how the parts of a whole are connected. Under basic mereological accounts, *Klaus' paw* is part of *Klaus* in much the same manner that *the caterpillar* is part of some abstract plural individual of *the caterpillar and the flower*. Yet, on some level, *Klaus* and *the caterpillar and the flower* are different. Klaus' paw is physically connected to the rest of Klaus' parts. The caterpillar is not connected to the flower. To capture the notions of how parts are connected to their wholes, a stronger theory is needed. Here enters `MEREOTOPOLOGY`, a theory which combines the intuitions of a part-whole relation with the topological notion of connectedness. As a field of study, mereotopology boomed in the turn of the century, with notable contributions to the field from Casati and Varzi (1999); Smith (1996); Varzi (1998).

[1]Correct when written on the 18th June 2024.

While mereotopology can describe types of wholes in terms of connectedness, it cannot capture all differences that are important to human perception and linguistic representations. Mereotopology is, for example, unable to describe the distance between disjoint parts of a whole. To capture that the caterpillar is touching the flower and not three centimetres away from the flower is not a matter of a mereology or mereotopology, but a mater of metric distance.

Similarly, capturing that the caterpillar is some solid whole and not a liquid whole is not necessarily a mereotopological concern, either. Mereotopology is concerned with parts, wholes, and connections, and not the constituent matter of the parts. Yet, the internal make-up of individuals is of perceptual and linguistic concern: humans have a prelinguistic substance/object distinction (Spelke and Kinzler, 2007), and languages have structures which are sensitive to the object/substance distinction such as stubbornly distributive predicates (Rothstein, 2010; Scwarzchild, 2011) and count comparison constructions (Bale and Barner, 2009; Barner and Snedeker, 2005).

When we talk about the metaphysical structure of individuals, it is clear we are dealing with a complex multifaceted story. With the background set, let us now turn to modelling mereological, mereotopological, substance/object, and distance concerns.

3.2 Algebraic concepts

To demonstrate the algebraic concepts that required for a theory of mereology and mereotopology, I introduce the basic concepts that underlie binary partial ordering and `LATTICE THEORY`. In what follows, all definitions are adapted from Grätzer (1998, 1-6), Partee et al. (1990, p.277-296) and Landman (2020, chapter 2).

The `PARTIAL ORDERING RELATION`, $\leq$, is generally understood as obeying the the axioms of `REFLEXIVITY, TRANSITIVITY` and `ANTISYMMETRY`.[2]

[2]I follow Varzi (2007) in simplifying notation in places by dropping universal quantifiers for reflexivity, transitivity and antisymmetry. These formulas should nonetheless to be understood as universally closed. See Wągiel (2021a) for versions with universal notation.

(1) `REFLEXIVITY`
x ≤ x

(2) `TRANSITIVITY`
(x ≤ y ∧ y ≤ z) → x ≤ z

(3) `ANTISYMMETRY`
(x ≤ y ∧ y ≤ x) → x = y

For any given set A, applying a partial order generates a partially ordered set, known also as a `POSET`. This is written as $\langle A, \leq \rangle$. Technically speaking, $\langle A, \leq \rangle$ is a subset of the Cartesian product of A. The type of poset that we will be concerned with are those where not all elements are `COMPARABLE`, *i.e.* for some x, y ∈ A, $\langle x, y \rangle \notin \langle A, \leq \rangle$, $\langle y, x \rangle \notin \langle A, \leq \rangle$. For example, set B in (5) is a partial order on A, as B is a subset of the Cartesian product of A and has the properties of reflexivity, antisymmetry, transitivity, and not all elements of B are comparable.

(4) $A = \{a, b, c, d\}$

(5) $B = \langle A, \leq \rangle = \{\langle a, a \rangle, \langle b, b \rangle, \langle c, c \rangle, \langle d, d \rangle, \langle a, b \rangle, \langle a, c \rangle, \langle a, d \rangle, \langle b, d \rangle, \langle c, d \rangle\}$

With this, it is useful to define `UPPER BOUND, LOWER BOUND, INFINITUM AND SUPREMUM`, and `LATTICE`. Consider an arbitrary poset P, and let $S \subseteq P$:

(6) `UPPER BOUND`
An upper bound of S is any element $a \in P$, if it exists, such that
$\forall b \in S \rightarrow b \leq a$

(7) `LEAST UPPER BOUND / SUPREMUM (SUP)`
Let T be the set of all upperbounds of S.
The least upper bound (SUP) of S is any element $x \in T$ such that
$\forall y \in T \rightarrow x \leq y$

(8) `LOWER BOUND`
A lower bound of S is any element $a \in P$, if it exists, such that
$\forall b \in S \rightarrow a \leq b$

(9) `GREATEST LOWER BOUND / INFINITUM (INF)`
Let Q be the set of all lower bounds of S.
The greatest lower bound (INF) of S is any element $x \in Q$ such that
$\forall y \in Q \rightarrow y \leq x$

(10) A poset $\langle P \leq \rangle$ is a `JOIN SEMI-LATTICE` iff the supremum exists for all $a, b \in P$

(11) A poset $\langle P \leq \rangle$ is a `MEET SEMI-LATTICE` iff the infinitum exists for all $a, b \in P$

(12) A poset $\langle P \leq \rangle$ is a `LATTICE` iff it is both a join-semi lattice and a meet semi-lattice, *i.e.* the supremum and the infinitum exist for all $a, b \in P$

Finally, given the definition of lattice, it is useful to define `JOIN` and `MEET`, which are both binary operations that apply to elements of a poset to yield an element of the poset. Both join and meet have the properties of `IDEMPOTENCY, COMMUTATIVITY, ASSOCIATIVITY` and `ABSORPTION`. Note, with these definitions set, it follows that if for any element $a \leq b$ then $a \wedge b = a$ and $a \vee b = b$.

(13) `MEET`
$a \wedge b \stackrel{\text{def}}{=} \textsc{inf}\{a,b\}$

(14) `JOIN`
$a \vee b \stackrel{\text{def}}{=} \textsc{sup}\{a,b\}$

(15) `IDEMPOTENCY`
$a \wedge a = a$
$a \vee a = a$

(16) COMMUTATIVITY
$a \wedge b = b \wedge a$
$a \vee b = b \vee a$

(17) ASSOCIATIVITY
$(a \wedge b) \wedge c = a \wedge (b \wedge c)$
$(a \vee b) \vee c = a \vee (b \vee c)$

(18) ABSORPTION
$a \wedge (a \vee b) = a$
$a \vee (a \wedge b) = a$

Posets are sets of ordered pairs, and so may be represented via general set notation, as in set B in (5), repeated below. It is common to represent posets via a HASSE DIAGRAM, where the lines (going upwards) represent the relation. The Hasse diagram in Figure 3.1 is equivalent to set B. Using a Hasse diagram, we can easily visualise the properties so far discussed. For example, it is clear to see that the properties of the partial order (reflexivity, antisymmetry, transitivity) are satisfied in the Hasse diagram in Figure 3.1. That is, the transitivity of *e.g.* $\{\langle a,b\rangle, \langle b,d\rangle, \langle a,d\rangle\}$ is visualised by the up-going lines. We can also see that the supremum and infinitum exist (d, a, respectively), and so A is a lattice. We can also describe d as the join of b and c ($d = b \vee c$), and also a as the meet of b and c ($a = b \wedge c$).

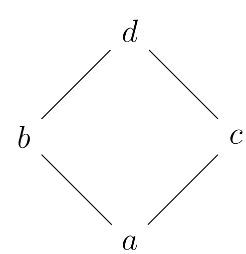


Figure 3.1: Hasse diagram of a lattice

(5) $B = \langle A, \leq\rangle = \{\langle a,a\rangle, \langle b,b\rangle, \langle c,c\rangle, \langle d,d\rangle, \langle a,b\rangle, \langle a,c\rangle, \langle a,d\rangle, \langle b,d\rangle, \langle c,d\rangle\}$

We now define `BOOLEAN ALGEBRAS` which have the unary operation `COMPLEMENT` and nullary operations 0 (bottom element, meet of all subsets), and 1 (top element, join of all subsets) defined. Given these operations along with join and meet, in a Boolean Algebra, the `TOP LAWS` and `BOTTOM LAWS` are obeyed. Figure 3.2 exemplifies a Boolean Algebra, where $1 = \textsc{sup}\{a, b, c, d\}$ and $0 = \textsc{inf}\{a, b, c, d\}$.

(19) `COMPLEMENTATION`
$a \wedge \neg a = 0$
$a \vee \neg a = 1$

(20) `TOP LAWS`
$a \wedge 1 = a$
$a \wedge 0 = 0$

(21) `BOTTOM LAWS`
$a \vee 0 = a$
$a \vee 1 = 1$

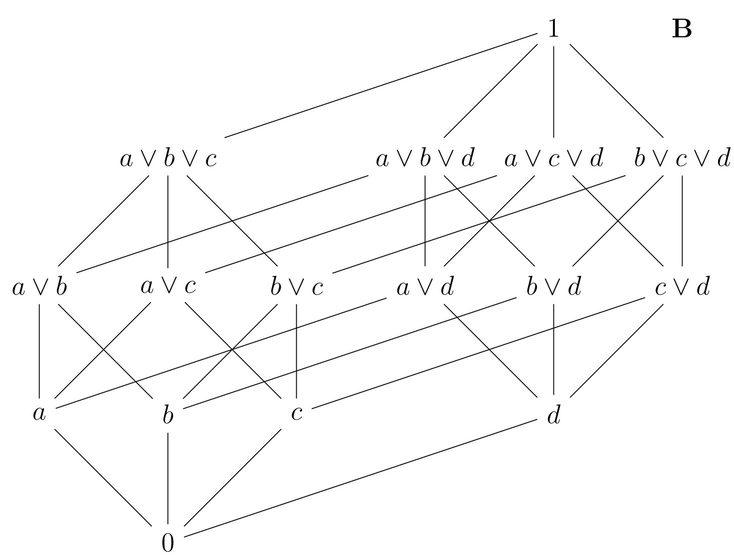


Figure 3.2: Hasse diagram of a Boolean Algebra

3.3 Mereology

With the basic algebraic concepts set, we now use these ideas to formalise part-whole relationships, *e.g.* for my cat Klaus' paw to be part of Klaus, but for the caterpillar to not be part of the flower he is sitting on. To discuss these relationships, we now build a ground mereology. The formalisation of mereology I assume uses standard definitions which I have variously adapted from Casati and Varzi (1999); Champollion (2010); Champollion and Krifka (2016); Grimm (2012b); Landman (2020); Varzi (2007, 2019) and Wągiel (2021b). As these definitions are standard, their presentation is also similar to other presentations in the literature. Although the presentation is not necessarily original, the concepts discussed are necessary perquisites for discussions later in this thesis.

Before settling into the discussion, I will clarify a terminological point. I use the term `INDIVIDUAL` pre-theoretically to refer to well defined objects/wholes that have privileged status in human perception/construals of the world. A cat, for example, is an individual. So too is a bikini. Half of a cat is not an individual in a pre-theoretical sense. In terms of mereology, however, the term `INDIVIDUAL` is used for any part or whole that is currently under discussion. In mereology, the half of a cat is an individual. In an attempt to avoid confusion between theoretical and non-theoretical uses, I follow Grimm (2012b) and Wągiel (2021b) in using the term `INDIVIDUAL` in its pre-theoretical use, and `M-INDIVIDUAL` for the mereological sense. In this way, every individual at the pre-theoretical level has a corresponding m-individual status that we can talk about. It is however not the case that every m-individual has a privileged status as an individual in a pre-theoretical sense: half a cat is very much an m-individual, but in some useful human sense, it is not necessarily an individual.

The most primitive relation in any formal mereological account is the `PART OF RELATION`, $\sqsubseteq$, which is modelled on intuitive parthood and is taken to be formalised as a partial order.[3] Being defined as a partial order, mereologies therefore obey the axioms of being reflexive, transitive, and antisymmetric, redefined here with respect to the $\sqsubseteq$ relation.

(22) `REFLEXIVITY`
$\mathrm{x} \sqsubseteq \mathrm{x}$
Everything is part of itself

(23) `TRANSITIVITY`
$(\mathrm{x} \sqsubseteq \mathrm{y} \wedge \mathrm{y} \sqsubseteq \mathrm{z}) \rightarrow \mathrm{x} \sqsubseteq \mathrm{z}$
If x is part of y, and y is part of z, then x is part of z

(24) `ANTISYMMETRY`
$(\mathrm{x} \sqsubseteq \mathrm{y} \wedge \mathrm{y} \sqsubseteq \mathrm{x}) \rightarrow \mathrm{x} = \mathrm{y}$
Two distinct things, x and y, cannot be part of each other

Considering the structure of matter as a mereology ordered by $\sqsubseteq$, it follows that the structure of matter is reflexive, transitive, and antisymmetric. For example, take my cat, Klaus. We may say that: Klaus is part of himself (reflexivity), Klaus' whiskers are part of his face, which are part of Klaus, ergo, Klaus' whiskers are part of Klaus (transitivity), and there is no distinct thing that is part of Klaus that Klaus is also part of (antisymmetry).

Alongside primitive $\sqsubseteq$ there are a number of axioms that are useful to define. First is the notion `PROPER PART`, $\sqsubset$, which is standardly defined as in (25)[4]. As is

[3]Throughout the literature, different authors have used varying symbols to represent the part of relation. For example, Champollion (2010); Champollion and Krifka (2016); Grimm (2012b) use the $\leq$ symbol, Landman (2020); Wągiel (2021b) use the $\sqsubseteq$ symbol, and Casati and Varzi (1999); Varzi (2019) simply use P, and similarly Rescher (1955) uses *Pt.* I choose to represent the partial order with the $\sqsubseteq$ symbol as to a.) distinguish from arbitrary partial orders which I used $\leq$ to represent in §3.2, and b.) follow Landman (2020), who's so-called `ICEBERG SEMANTICS` which I will adapt later from chapter 5, and c.) avoid confusion with later use of distance functions which use $\leq$ in a standard mathematical sense *i.e.*'less than or equal to'.

[4]*c.f.* Champollion (2010) who defines proper part as $\mathrm{x} \sqsubset \mathrm{y} \overset{\mathsf{def}}{=} \mathrm{x} \sqsubseteq \mathrm{y} \wedge \mathrm{x} \neq \mathrm{y}$ (adapted), which is equivalent to (25).

the case with the $\sqsubseteq$ relation, the $\sqsubset$ relation is transitive and antisymmetric. However, $\sqsubset$ is not reflexive (nothing is a proper part of itself). Intuitively, with this definition we may say that Klaus' paw is a proper part of Klaus (*Klaus's-paw*$\sqsubset$*Klaus* ✓), but we cannot say that Klaus is a proper part of Klaus (*Klaus*$\sqsubset$*Klaus* ✗).

(25) PROPER PART, $\sqsubset$
$x \sqsubset y \stackrel{\text{def}}{=} x \sqsubseteq y \wedge \neg(y \sqsubseteq x)$
x is a proper part of y if x is part of y, and y is not a part of x

We may also want to talk of m-individuals which are not part of each other, but which overlap, *i.e.* they share parts. The ancillary relation OVERLAP is defined in (26)[5]. The converse, *i.e.* where elements share no parts, is DISJOINTEDNESS[6], as defined in (27).[7] Examples of overlap include a glass of water split into thirds. The bottom two-thirds and the top two-thirds of water in the glass overlap, as they share a part: the middle third. Conversely, the top-third of the water in this example is disjoint from the bottom third, as they share no parts.

(26) OVERLAP, $\mathcal{O}$
$\mathcal{O}(x,y) \stackrel{\text{def}}{=} \exists z[z \sqsubseteq x \wedge z \sqsubseteq y]$
x and y overlap if there is something that is part of both

(27) DISJOINT, $\mathcal{D}$
$\mathcal{D}(x,y) \stackrel{\text{def}}{=} \neg\mathcal{O}(x,y)$
x and y are disjoint if they don't overlap

[5]Again, symbols in the literature differ for overlap. Champollion (2010) uses $x \otimes y$, Wągiel (2021b) uses lowercase o(x,y), Casati and Varzi (1999); Grimm (2012b) use uppercase O(xy), and Champollion and Krifka (2016) use $x \circ y$.

[6]Overlap and disjointedness are formalised slightly differently in (Landman, 2020, p.18), where overlap is defined as the case where two elements share a common meet (that is not 0, as the bottom element is removed).

[7]Disjoint m-individuals are sometimes called DISCRETE, as in Rescher (1955) and Smith (1996).

Considering that $\sqsubseteq$ is only defined to be a partial order, there is nothing in the model so far that rules out structures that are unintuitive and not in the spirit of what a part-whole relationship should capture. For example, the structure in Figure 3.3 allows a to be a proper part of b, all while b has no other proper parts. For a theory of mereology which intends to capture intuitive parts and wholes, this is unwanted, as if anything has proper parts then it must have multiple proper parts. Additionally, the structure in Figure 3.4 is unwanted in a mereological theory. This structure shows two distinct elements, c and d, share the same parts. Under a part-whole view, we would want to say that if two m-individuals share the same parts and have no other parts, they are the same m-individual. To rule out structures such as Figure 3.3 and Figure 3.4, mereologies must have additional conditions placed on the $\sqsubseteq$ relation.

$$
\begin{array}{c}
b \\
| \\
a
\end{array}
$$

Figure 3.3: An m-individual which has only one proper part

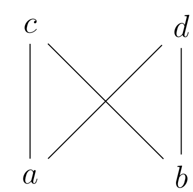


Figure 3.4: Two distinct m-individuals share the same parts

The first condition needed is `SUPPLEMENTATION` as defined in (28). This rules out structures which have an element which has only one proper part, *i.e.* structure Figure 3.3.[8]

(28) `SUPPLEMENTATION`
$x \sqsubset y \rightarrow \exists z[z \sqsubseteq y \wedge \neg\mathcal{O}(z,x)]$
If x is a proper part of y, then there is another proper part of y that does not overlap with x

[8]As it turns out, supplementation is denied in some theories of mereology. I will not explore these options here, but see Simons (1991) for a brief overview of the history of mereology in the 1900s.

To rule out structures such as in Figure 3.4, `SUM` and `UNIQUENESS OF SUMS` are required. Consider the definition of `SUM` in (29), which defines an x as the join of all overlapping elements of a set P. In classical mereology, sums are required to be unique. This is defined in (30). The uniqueness of sums rules out the structure in Figure 3.4 as *c* and *d* are both sums with identical parts, meaning that these sums are not unique, violating (30).[9]

(29) `SUM`
$\mathrm{sum}(x,P) \stackrel{\text{def}}{=} \forall y[P(y) \rightarrow y \sqsubseteq x] \wedge \forall z[z \sqsubseteq x \rightarrow \exists z'[P(z') \wedge \mathcal{O}(z,z')]]$
x is the sum of things in a set P if: everything in P is a (not necessarily proper) part of x, and for everything that is part of x there is something distinct which overlaps with that thing

(30) `UNIQUENESS OF SUMS`
$\forall P[P \neq \varnothing \rightarrow \exists! z\ \mathrm{sum}(z,P)]$
Every non-empty set has a unique sum

Since Tarski (1935), it is known that placing these conditions upon $\sqsubseteq$, the mereological structure created is a Boolean algebra without the bottom element (0), *i.e.* a Boolean semi-lattice (the join of all subsets is defined). This is exemplified in Figure 3.5. Explicitly, we can see that the mereology in Figure 3.5 is `SUPPLEMENTED`, as every element that has a proper part has at least two proper parts. Further, the mereology in Figure 3.5 is unique as there are no two distinct sums which share identical proper parts.

[9]The formalism for uniqueness of sums is taken directly from Champollion and Krifka (2016, p.517) and Wągiel (2021b, p.187).

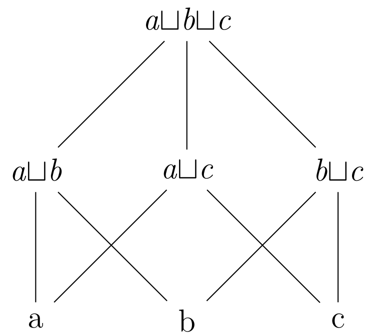


Figure 3.5: Mereology as a Boolean semi-lattice

It is useful to define the further notions of `BINARY SUM` and `GENERALISED SUM`. The binary sum operation is useful in that given any two elements of a set, the output will be the least upper bound of those two elements. For example, the binary sum of a and b is $a \sqcup b$. As binary sum maps onto the notions of `JOIN` defined for an arbitrary partially ordered set it has the properties of idempotency, commutativity, associativity and absorption.[10] Therefore, the binary sum of $a \sqcup b$ and $a \sqcup c$ is $a \sqcup b \sqcup c$. In the mereology in Figure 3.5, as uniqueness of sums is defined, binary sum holds for any two elements of the mereology. Following this, the notion of `GENERALISED SUM` is useful, which sums the elements of a given set which has an arbitrary number of elements.

(31) `BINARY SUM`
$\mathrm{x} \sqcup \mathrm{y} \stackrel{\text{def}}{=} \iota \mathrm{z}\ \mathrm{sum}(\mathrm{z},\{\mathrm{x},\mathrm{y}\})$

(32) `GENERALISED SUM`
$\sqcup \mathrm{X} \stackrel{\text{def}}{=} \mathrm{sum}(\mathrm{z},\mathrm{X})$, where X is any non-empty set

We can now define `ATOM` and `ATOMICITY`. Simply put, an atom is any element in a given mereology that has no proper parts. Following this, if any mereology is

[10]In many works that use mereological (and mereotopological) formalisations, there is free interchangeable use of `SUM` and `JOIN`.

ultimately made up of atoms, it may be described as `ATOMIC`, or that it has the property of `ATOMICITY`. In other words, if every m-individual has atoms as parts, then the mereology is atomic. The mereology in Figure 3.5 is atomic, as a, b and c are atoms (they have no proper parts). This makes $\{a, b, c\}$ (the lowest tier of the semi-lattice) the atomic set of the mereology in Figure 3.5.

(33) `ATOM`

$$\text{atom(x)} \stackrel{\text{def}}{=} \neg\exists\text{y}[\text{y} \sqsubset \text{x}]$$

x is an atom if it has no proper parts

(34) `ATOMICITY`

$$\forall\text{y}\exists\text{x}[\text{atom(x)} \wedge \text{x} \sqsubseteq \text{y}]$$

With this, we now have a `GROUND MEREOLOGY`.

3.4 Mereotopology

Though mereology has been immensely successful for describing *part-of* relations, it is not without its drawbacks. One criticism of standard mereology is that transitivity is argued to not hold in all *part-of* relations. For example, Rescher (1955) argues that:

> A part (*i.e.*, biological subunit) of a cell is not said to be a part of the organ of which that cell is a part.
> (Rescher, 1955, p.10)

In the linguistic sphere, Moltmann (1998) has called transitivity into question for similar reasons. However, the use of common day senses of ‘PART’ to define $\sqsubseteq$ is generally problematic, as PART has many senses, an observation that goes back to Aristotle’s *Metaphysics*, book Δ. The major criticism of relying on natural language PART to define $\sqsubseteq$ is discussed explicitly by Casati and Varzi (1999, p33-34), who state that Rescher and Moltmann's intuitions rely on a sense of PART that is narrowed: it requires all parts of a whole to hold the same functions of a whole, a condition not to be placed upon $\sqsubseteq$. Similarly, Champollion (2010); Champollion and Krifka (2016) discuss parallel issues relate to `STRUCTURED` and `UNSTRUCTURED` parthood. Unstructured parthood is the equivalent of arbitrary sum formation, while structured parthood is the relationship between an m-individual and its parts. The example Champollion (2010, p.11) gives is that of a ham sandwich; as a structured whole, the ham sandwich is more than a sum of some ham and some bread. For Champollion and Champollion and Krifka, different aspects of natural language map onto different types of parthood. Unstructured parthood maps to the relationship between singular individuals and their plurals, while structured parthood corresponds to the relationship between an individual and its parts. Champollion points out that the common use of PART may concern itself with structured parthood, or it may not, but the point is that the the notion of $\sqsubseteq$ does not necessarily map onto the natural language use of PART, and the formulation of $\sqsubseteq$ should remain separate.

There is one other major criticism of standard mereology that is not rooted in analysis of $\sqsubseteq$ being equivalent to natural language PART. Namely, the inability for standard mereology to capture something that is intuitively obvious: the distinction between, as Casati and Varzi (1999, p.51) put it, 'good' and 'bad' wholes, as well as a 'real' part-whole relationship. The problems that classical mereology faces in these areas are due to the unrestricted sum formation, *i.e.* if two m-individuals exist, so does their binary sum. Therefore, the binary sums listed in (35) are equally legal in mereology with no discernible difference between the status of the sum.

(35) Legal binary sums in mereology:

a. *my-cat*$\sqcup$*my-duck*

b. two halves of a sphere

c. *my-fingers*$\sqcup$*my-palm*

d. *my-fingers*$\sqcup$*Edward's-fingers*

The sums in (35) are intuitively different one some real-world level. The sum of *my-cat*$\sqcup$*my-duck* is abstract as my cat and my duck are not in any way physically connected. In a very real way, the two m-individuals, *my cat* and *my duck* are disjoint with space between them, and yet under mereological assumptions they are part of ($\sqsubseteq$) the sum which only exists due to the unlimited sum formation. The sum of two parts of a sphere on the other hand is not abstract in the same manner. Rather, the two halves of a sphere form an integrated whole in its own right; the two halves of the sphere are connected in a way that there is no space between them. The problem holds for as many partitions of potential sums imaginable. The sum of *my-fingers*$\sqcup$*my-palm* is a useful whole, so much so that there is a term for it: MY HAND. Yet, from a mereological point of view, a sum is all there is. It is no different, really, than the sum of *my-fingers*$\sqcup$*Edward's-fingers*, as they are both just sums. The fact that my fingers are connected to my palm in such a way that I cannot remove them (without considerable pain) is wholly irrelevant under standard mereology. So, while connection and parthood relations are intuitively different, mereology fails to capture these differences.

In the philosophical sphere, an early attempt to capture the intuition between arbitrary and non-arbitrary wholes is found in Whitehead (1920, 1929), though the focus is on the ontology of events. In more recent philosophical work there have been moves to capture different part-whole distinctions in the ontology of spatial entities (Casati and Varzi, 1999; Smith, 1996; Varzi, 2007, 2019). This is done by incorporating topological notions into mereology. This is known as `MEREOTOPOLOGY`. Like the discussion of standard mereology above, as the definitions of mereotopology here are fairly standard, and so their presentation is also similar to other presentations in the literature. In the following, I follow closely the version of mereotopology as it is presented in Casati and Varzi (1999); Varzi (2007), though notation is adapted and the ideas are presented in a way heavily influenced by both Grimm (2012b) and Wągiel (2021b).

In topology, the crucial notion is `CONNECTEDNESS`, $\mathcal{C}$, which is taken to be primitive. The connectedness relation is reflexive and symmetric. $\mathcal{C}$ differs from

$\sqsubseteq$ in its axioms, particularly in that $\mathcal{C}$ is not required to be transitive. This allows an m-individual a to be connected to an m-individual b which is connected to an m-individual c without having to require a and c to be connected. The example that Casati and Varzi (1999, p.52) give puts this idea of connectedness concretely in terms of fiat connectedness: France is connected to Italy and Italy to Slovenia, but France and Slovenia are disconnected.[11]

(36) `REFLEXIVITY`
$\mathcal{C}$(x,x)
Everything is connected to itself

(37) `SYMMETRY`
$\mathcal{C}$(x,y) $\rightarrow$ $\mathcal{C}$(y,x)
If x is connected to y, then y is also connected to x

The enrichment of mereology with $\mathcal{C}$ is quite simple, as the $\sqsubseteq$ relation and the topological $\mathcal{C}$ relation 'speak' to each other. If a given m-individual x is part of a given m-individual y, it is intuitively obvious that x and y are connected. To formalise this, the bridging principle `MONOTONICITY` is defined, creating the ground theory of mereotopology, which implies that mereological overlap is a form of connection.[12] The axiom of monotonicity implies both `INTEGRITY` and `UNITY`.

(38) `MONOTONICITY`
x $\sqsubseteq$ y $\rightarrow$ $\forall$z[$\mathcal{C}$(z,x) $\rightarrow$ $\mathcal{C}$(z,y)]
If x is part of y, then anything that is connected to x is also connected to y

(39) `INTEGRITY`
$\mathcal{O}$(x,y) $\rightarrow$ $\mathcal{C}$(x,y)
If x overlaps y, then x is connected to y

[11] Again, I follow Varzi (2007) in simplifying notation by dropping universal quantifiers for reflexivity symmetry. These formulas should nonetheless to be understood as universally closed. See Wągiel (2021a) for versions with universal notation.

[12] The axiom of monotonicity actually makes the topological axiom of symmetry redundant.

(40) `UNITY`
$x \sqsubseteq y \rightarrow \mathcal{C}(x,y)$
If x is part of y, then x is connected to y

We have a ground mereotopology. We are now in a position to define different types of overlap, such as `INTERNAL PART`, `INTERNAL OVERLAP`, and `TANGENTIAL OVERLAP`.[13] The Internal Part relation is shown visually in Figure 3.6, where *b* is an internal part of *a*. The Internal Overlap relation is shown in Figure 3.7, where only a part of *b* overlaps with *a*. Finally, Figure 3.8 exemplifies Tangential Overlap, where *b* and *a* overlap but their interiors do not overlap.

(41) `INTERNAL PART`
$\textsc{ip}(x,y) \stackrel{\text{def}}{=} x \sqsubseteq y \wedge \forall z[\mathcal{C}(z,x) \rightarrow \mathcal{O}(z,y)]$
x is an internal part of y if x is part of y, and everything that is connected to x overlaps y

(42) `INTERNAL OVERLAP`
$\textsc{io}(x,y) \stackrel{\text{def}}{=} \exists z[\textsc{ip}(z,x) \wedge \textsc{ip}(z,y)]$
x internally overlaps y if there is an internal part of x that is also an internal part of y

(43) `TANGENTIAL OVERLAP`
$\textsc{to}(x,y) \stackrel{\text{def}}{=} \mathcal{O}(x,y) \wedge \neg\textsc{io}(x,y)$
x and y tangentially overlap if x and y overlap, but their interiors do not overlap

[13] For transparency, Casati and Varzi (1999, p.55) also defines other connection relations that I do not discuss here.

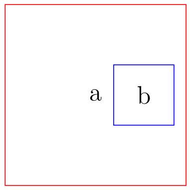


Figure 3.6: Internal part

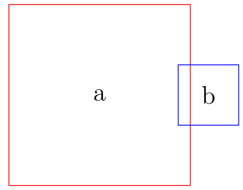


Figure 3.7: Internal overlap

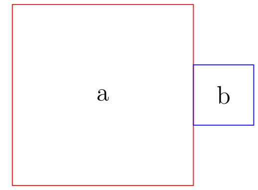


Figure 3.8: Tangential overlap

We now define `INTERIOR`[14], `EXTERIOR`, `CLOSURE` and `BOUNDARY`. The interior of an m-individual x can be described as the sum of all its internal parts. This is exemplified in Figure 3.9, where the shaded area is the interior of m-individual a. From this, the exterior is defined as the complement of the interior, exemplified in Figure 3.10. Next, `CLOSURE` is defined as the complement of the exterior of x, which captures both the interior and any boundaries, exemplified in Figure 3.11. Finally, we may now define `BOUNDARY` as the complement of the sum of the interior and exterior of an m-individual x.

(44) `INTERIOR`
$\textsc{int}(\mathrm{x}) \stackrel{\text{def}}{=} \sqcup \mathrm{X}$ where $\mathrm{X} = \{\mathrm{y}\text{: } \textsc{ip}(\mathrm{y},\mathrm{x}) = \mathrm{TRUE}\}$

(45) `EXTERIOR`
$\textsc{ext}(\mathrm{x}) \stackrel{\text{def}}{=} \textsc{int}(\neg(\mathrm{x}))$

(46) `CLOSURE`
$\textsc{clo}(\mathrm{x}) \stackrel{\text{def}}{=} \neg(\textsc{ext}(\mathrm{x}))$

(47) `BOUNDARY`
$\textsc{bou}(\mathrm{x}) \stackrel{\text{def}}{=} \neg(\textsc{int}(\mathrm{x}) \sqcup \textsc{ext}(\mathrm{x}))$

[14]This definition of `INTERIOR` is specifically adapted from Wągiel (2021b, p. 195), which is adapted from Casati and Varzi (1999, p. 45).

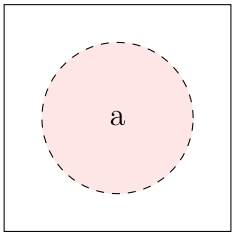


Figure 3.9: Interior

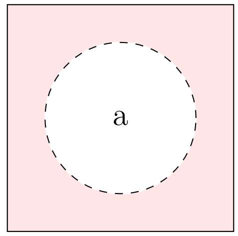


Figure 3.10: Exterior

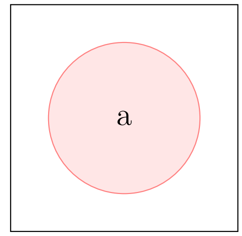


Figure 3.11: Closure

It is now possible to start building a definition of what it means to be an `INTEGRATED WHOLE`. Such a definition is useful, as it enables us to define the difference between wholes which come as one, such as a cat, and wholes which come as scattered pieces, such as a bikini. The first step to this goal is to define a `SELF CONNECTED` whole, which is any whole that cannot be split into two disconnected parts, as in (48). In this way, the sum *bikini-bottom*$\sqcup$*bikini-top* is ill defined for self connectedness, though the sum of two halves of a sphere, $a \sqcup b$, is well defined for self connectedness.

(48) `SELF CONNECTED`
$\textsc{sc}(x) \stackrel{\text{def}}{=} \forall y \forall z[\forall w[\mathcal{O}(w,x) \leftrightarrow (\mathcal{O}(w,y) \vee \mathcal{O}(w,z))] \rightarrow \mathcal{C}(y,z)]$
An m-individual x is self-connected if any two parts that make up the whole of x are connected to each other

The property of being self connected is not enough to capture the fully desired notion of an integrated whole, as it does not rule out sums where the parts merely 'touch' at a single point, such as the sum of the two spheres $a \sqcup b$ in Figure 3.12, where the boundaries of a and b are connected at a minimal point. To obtain a stronger notion of an integrated whole, the property of `STRONGLY SELF CONNECTED` in (49) is defined, where it is not only the sum that is self-connected, but so too is its interior. Such a definition rules out the sum of $a \sqcup b$ in Figure 3.12, as the interiors of $a \sqcup b$ are not defined for self connectedness.

(49) STRONGLY SELF CONNECTED

$\textsc{ssc}(x) \stackrel{\text{def}}{=} \textsc{sc}(x) \wedge \textsc{sc}(\textsc{int}(x))$

x is strongly self connected if it is self connected and its interior is self-connected.

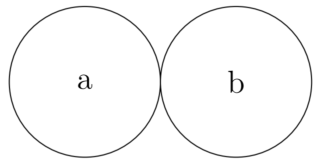


Figure 3.12: Self connectedness is defined, strongly self connectedness is not defined

With the definition of strongly self connected, we have a vast improvement toward the goal of describing an integrated whole. Yet, strongly self connectedness alone is not sufficient. The example that Casati and Varzi (1999) give is that of a ball. The bottom half of a ball counts as strongly self connected, but in and of itself is not an integrated whole in the way that the whole ball is. To define what it means to be an integrated whole, the property of MAXIMALLY STRONGLY SELF CONNECTED (MSSC) is required. An MSSC whole is any strongly self connected sum that has any overlapping strongly self connected parts as part of the whole sum. The property of being maximally strongly self connected then stands relative to a property, as in (51), where ϕ stands for the property in question.[15]

(50) MAXIMALLY STRONGLY SELF CONNECTED

$\textsc{mssc}(x) \stackrel{\text{def}}{=} \textsc{ssc}(x) \wedge \forall y\ [\textsc{ssc}(y) \wedge \mathcal{O}(y,x) \rightarrow y \sqsubseteq x]$

An m-individual is maximally strongly self-connected if (i) every part of the m-individual is connected to (overlaps) the whole (strongly self-connected) and (ii) anything else which overlaps it and is strongly self-connected is once again part of it (maximality).)

[15] The translations for (50) and (51) are taken directly from Wągiel (2021b, p.198).

(51) MAXIMALLY STRONGLY SELF CONNECTED WITH RESPECT TO A PROPERTY
$\textsc{mssc}(x,P) \stackrel{\text{def}}{=} P(x) \wedge \textsc{ssc}(x) \wedge \forall y[P(y) \wedge \textsc{ssc}(y) \wedge \mathcal{O}(y,x) \rightarrow y \sqsubseteq x]$
An m-individual is maximally strongly self connected relative to a property if (i) every part of the m-individual is connected (overlaps) the whole (strongly self connected) and (ii) anything which has the same property, is strongly self-connected, and overlaps it is once again part of it (maximality).

The development of mereotopology allows for definitions of distinct types of sums that are not available in standard mereology. The sum of *my-duck*$\sqcup$*my-dog* in (35) can be described via the standard mereological sum operator. It is simply the (abstract) join of two distinct and disjoint m-individuals. If it is a whole, it is a whole which is *not* self connected. The sum of two halves of a sphere on the other hand is not an arbitrary sum, but a sum that is defined for being MSSC, so long as the halves are not separated.

3.5 Substances and non-substances

With the tools of mereotopology, we now have a working definition of an integrated whole as an MSSC m-individual with respect to a property. We do not have a definition that distinguishes internal make-up of those wholes. While the property of being MSSC may hold for integrated wholes such as a ball, a cat, or the caterpillar on the flower, what can we say of substances such as the water in my glass? As it turns out, substances (in the naïve sense) may themselves be described as MSSC wholes. The water in my glass may very well count as an MSSC m-individual so long as any water that is connected to the water in my glass is part of the water in my glass. In standard mereotopological accounts, there is no discussion of substance verses non-substance. And why should there be? Mereotopology does not care for the composition of matter, but only for connectedness and parthood. Can we do better? Can we define, in mereotopological terms, the difference between substances like water and non-substances like cats?

As it turns out, capturing such a distinction is tricky. There is a delicate balance between the description of an m-individual in a given context, and the construals of properties. For one, construed standards of individuation (relative to a property) are not reducible to any one fundamental mereotopological concept. Both the properties of CAT and BIKINI have a standard for what counts as one cat or one bikini in any given context. Any given cat in context is an integrated whole, *i.e.* an MSSC whole with respect to the cat-property. A given bikini on the other hand is not an MSSC m-individual, but rather a scattered whole that fails to meet the criteria for self connectedness.

Construed standards of individuation for a given property are made on a human conceptual level, rather than any fundamental level. What seems to be important for standards of individuation at the human level is some judgement of `FUNCTION`.[16] Consider that in a theory of mereotopology *the Eiffel tower and my cat Klaus* are as good a scattered whole as *a bikini top and a bikini bottom* as both are sums of disjoint MSSC m-individuals. However, in a real human sense, *the-bikini-top*$\sqcup$*the-bikini-bottom* is an intuitively *useful* scattered whole in a way in which *Klaus*$\sqcup$*The-Eiffel-Tower* is not. The usefulness of one scattered whole and the none usefulness of another is ultimately not a matter of mereotopology, but is down to human evaluation. The difference in the usefulness evaluation has, I surmise, resulted in terms for useful properties in natural language. In this case, BIKINI is a label for a useful scattered whole, but there is no label such as CATTOWER for the sum of *Klaus*$\sqcup$*The-Eiffel-Tower*. In theory, there could be a label CATTOWER which contains information about a useful scattered whole of cats and metal towers, but this label, I presume, does not exist in any natural language.

With this, I suggest that an `INDIVIDUAL` in a non-mereological, pre-theoretical, sense can be described using mereotopological means such that an individual is any m-individual that has privileged status at a conceptual human level, be it an MSSC whole, scattered whole, or any other relevant type of whole. The definition in (52)

[16]This reasoning is comparable to Landman (2020)'s discussion of `VERTICAL OVERLAP` as a litmus for individuation.

captures MSSC wholes such as cats and scattered wholes such as bikinis. In theory, (52) would capture scattered wholes such as *Klaus*$\sqcup$*The-Eiffel-Tower* only if there is an associated property P that privileges such a sum, which I assume there is not.

(52) INDIVIDUAL(x) $\stackrel{\text{def}}{=}$
$\forall$P[P(x) $\rightarrow$ (MSSC(x,P) $\vee$ SCATTER(x,P) ...)]
x is an individual if. for any property P, x is maximally strongly self connected relative to property P, or is a scattered whole relative to property P ...

The definition of INDIVIDUAL(x) in (52) does not exclude reference to m-individuals we may think of as liquids. All that (52) does is separate arbitrary sums from privileged sums in some real human sense. This is probably correct, as liquids such as water can have MSSC instantiations in a given context. There must therefore be some other definition that captures the fundamental difference between substances and non-substances.

One candidate to capture the difference between substances and non-substances is the behaviour of `BOUNDARIES`. Borrowing Casati and Varzi (1999, p.87-88)'s discussion of 'topological catastrophes', consider a drop of oil.[17] This drop of oil is an MSSC m-individual with respect to the property OIL, which has its own boundaries where the whole starts, and where the whole ends. Splitting the drop of oil into two separate parts will result in two distinct MSSC m-individuals with respect to the property OIL, each with their own boundaries. We may undo our splitting by moving these distinct oil m-individuals into contact so that the boundaries collide such that the two drops become once again a single whole, *i.e.* an MSSC individual with respect to the property OIL. This is extendable to any liquid; we may collate liquids into larger and larger MSSC m-individuals with the same property. In other words, for liquids we can impose absolute topological change with ease. In a non-technical sense, the boundaries of a liquid are completely malleable, collapsible and absorbable.

[17]For clarity, Casati and Varzi (1999)'s discussion of topological catastrophes is done so in the realm of boundaries, and not the substance/object distinction. I borrow the discussion here to exemplify how mereotopology, as presented, is not well equipped to deal with the substance/object distinction.

Now consider the same experiment, but with a cat. Imposing absolute topological change is quite different. Splitting a cat into disjoint halves will reveal two m-individuals, but it is unclear whether these m-individuals are MSSC with respect to the property CAT. Probably not. Similarly, if we take two cats, no matter how hard we push them together, there will be no absolute topological change in the boundaries such that we will end up with one cat. Two cats simply cannot become one cat. This lack of amalgamation for the m-individuals is not a facet of language, but rather some physical restriction of the world which does not clearly hold for *e.g.* oil m-individuals.

Consider our experiment again, but now with non-liquid materials of matter, *e.g.* gold, or wood. Attempting to impose topological change on non-liquid materials of matter is again different. Like liquids, splitting any MSSC hunk of gold will reveal two distinct MSSC m-individuals, complete with new boundaries. However, like with cats, try as I might, pressing two hunks of gold together will not result in absolute topological change (at least, not easily without employing drastic means such as using a crucible).

Our thought experiment shows that on some intuitive level the material constituent gold and the material constituent of water are akin. They are both made up of naïvely homogenous *stuff*. This is captured by the unifying feature of permissible absolute topological change in the *division* of m-individuals into disjoint m-individuals of the same property. However, it is well known that a divisional property is not a good litmus for what it means to be a substance or stuff (see Doetjes, 2021; Pelletier, 2012; Rothstein, 2010; Quine, 1960, *a.o.*). That is, while water, oil, or gold may be divided into instances of water, oil, and gold, talk of division introduces issues regarding infinite division. We can divide 500ml of water into two 250ml of water. We can then divide one of the 250ml portions into a 50ml and 200ml portion. We can keep dividing our water into increasingly smaller portions. There however comes a point where you can divide no further. Perhaps this point of no further division is when you have divided into hydrogen and oxygen molecules. The point is, a divisional property for substance/stuffhood is problematic as infinite division cannot be correct as there are ultimately proper parts of *e.g.* water that are not water.

So, division of m-individuals which maintain the property in question is not an entirely robust measure of substance/stuffhood. Of course, we could assume along with Bunt (1985); Chierchia (2010b) and Moltmann (2020) that there is a distinction between levels of ontology. Perhaps what is perceived is abstracted from what is fundamental (which may also be different to what is linguistically represented). However, how is this abstraction modelled? Pelletier (2012) points out that people have mental understandings that water, on some level, has smallest parts. Does knowledge of the molecular level play into construals of substance and object? Evidently not. Water at the molecular level hardly impacts our daily interactions with the stuff: we still consider water to be *stuff* in the way do not consider a cat as *stuff*. The issue of infinite division of water in daily life does not arise, and it hardly affects our construals of the `WATER` property. It would seem that abstracting what is perceived/construed from what is fundamental is obviously in some way correct, but it is not obvious how this should be achieved via a divisional property.

As it turns out, there is a mereotopological concept that we can borrow to capture the intuitions of division in a way that is both non-infinite and captures perceptual realities. I follow Grimm (2012a,b) and Wągiel (2021b) in suggesting that the mereotopological property `FIRM CONNECTION` is a useful starting point that captures the behaviour of substances/stuff. Explicitly, firm connection holds when two m-individuals have parts which have an SSC sum.

(53) `FIRMLY CONNECTED`

$\textsc{fc}(x,y) \overset{\text{def}}{=} \exists w \exists z[w \sqsubseteq x \wedge z \sqsubseteq y \wedge \textsc{ssc}(w \sqcup z)]$

Two m-individuals x and y are firmly connected if a sum of their parts is strongly self connected.

Both Grimm and Wągiel suggest that firm connection can be used to describe the behaviour of substances. The reasoning, originally formulated by Grimm (2012b), is that for a given substance, any proper part of that substance is firmly connected to another instance of that substance. Such a property clearly holds for water and gold, but does not clearly hold for an integrated whole, such as a cat or a scattered whole, such as a bikini. This being the case, firm connection as it stands in (53) does not actually describe what it means to be a substance/stuff. Two non-separated halves of a cat are firmly connected m-individuals, but clearly do not count as a substance. Therefore, to define `SUBSTANCE`, firm connection must stand relative to a property as in (54).

(54) $\textsc{substance}(x) \overset{\text{def}}{=}$
$\forall P[P(x) \rightarrow \exists y[P(y) \wedge x \neq y \wedge \textsc{fc}(x,y)]$
x is a substance if. for any property P x is firmly connected to a distinct m-individual of the same property.

The use of firm connection is useful for the description of a substance, as it privileges the material composition of an m-individual, but without commitment to infinite division or topological catastrophes. The fact that the boundaries of water (but not gold) can collapse or collate is not relevant for the description of a substance, nor is division. Rather, the difference lay in whether the material composition of the m-individual is of the same property as the whole. With this, the fundamental difference between objects and substances is the property of firm connectedness, where substances are m-individuals which are firmly connected to another m-individual of the same property. In essence, a substance/stuff judgment is made based on the perspective of the constituent matter as defined by firm connection.

3.6 Distance

Mereo(topo)logy is a theory of parts and wholes. We may be able to say that a caterpillar is an MSSC individual, as is the flower it is sitting upon. We may also say that *the-caterpillar*$\sqcup$*the-flower* is an arbitrary sum. But under mereotopology we can say nothing more than this. There are no tools in mereotopology to describe the distance between the caterpillar and the flower. Even if the caterpillar is sitting on the flower, with a vanishingly small distance between them, the flower and the caterpillar share no boundaries nor parts. Any connection relation is inappropriate to describe the distance, however small, between them. The issue of touch not being a mereotopological connection issue is put succinctly by Casati and Varzi:

> Thus, the book is on the table, but these two objects are simply not in touch if this is understood in terms of external connection. This is in agreement with physics and with ordinary topology....
> ... Can we also do justice to the common-sense intuition that somehow the book and the table are in touch—that nothing can be squeezed in between them? The proper answer is that this is not a mereotopological problem, but a metric one
> (Casati and Varzi, 1999, p.94)

As the nature of the distance between the caterpillar and the flower or the book and the table is not a mereotopological concern, but a distance concern, further primitives must be defined to capture this notion. To do so, consider a `DISTANCE FUNCTION`, d, which takes two individuals as arguments and outputs a Euclidean measure of their distance. For example, if x is our caterpillar, and y is our flower, and x and y are 3 metres away from each other, plugging these individuals into d will output 3 metres. Formally, the distance function is from the Cartesian product of a set with itself into real numbers: d: S $\times$ S $\rightarrow$ R. Assuming that x,y,z $\in$ S, the axioms for d are listed in (55).

(55) `STANDARD AXIOMS OF DISTANCE`, d

d(x,y) = 0 iff x=y

d(x,y) > 0

d(x,y) = d(y,x)

d(x,z) $\leq$ d(x,y) + d(y,z)

3.7 Summary and next steps

This chapter has presented almost no discussion of language, but rather introduced the theory of mereotopology, which is more robust than its predecessor mereology in capturing differences between different types of wholes. Mereotopology allows a useful definition of an integrated whole as an individual that is defined for MSSC status. This chapter has also outlined further useful notions, such as a basis for a fundamental distinction between substances and objects via manipulation of the firm connection relation. This chapter has also been explicit regarding that there is no role of distance in standard mereotopology. Rather, distance is taken to be a purely metric phenomenon, and is treated as its own primitive notion.

The assumptions laid out in this chapter will be useful in the forthcoming discussion. Specifically, the foundations of mereology, $\sqsubseteq$ and $\sqcup$, will be used to model a formal theory of number in line with formal semantic tradition dating back to Link (1983). Further, distance functions will be used in chapter 5 to characterise collective/singulative systems. Finally, the discussion of substance and non-substance will be relevant for chapter 7, where I discuss nominal flexibility.

Chapter 4

Construed Ontologies and Linguistic Ontologies

This chapter offers a discussion of two approaches to modelling number from the perspective of collective/singulative languages: the `LEXICALIST` approach (number features are generated in the lexicon) and `CONSTRUCTIONIST` approach (number features are derived in the syntax). I claim that both approaches have strengths in complementary areas. The lexicalist approach, specifically the `LEXICO-MEANING` approach, shines in its ability to weave construed ontologies and linguistic ontologies in such a way that construals of the world influence distribution of number neutral uncountable terms in and across language. However, lexico-meaning approaches struggle to capture crosslinguistic variation, where proposed mappings from perceptual realities to linguistic representations cannot capture variation without resorting to unfalsifiable means, such as arbitrariness and cultural divergence. Contrastively, the constructionist approach has powerful tools to account for crosslinguistic variation, but suffers in that the system has no room to incorporate the very real ways in which perceptual realities influence linguistic systems. Ultimately, I claim that a fusion of the strengths of the lexico-meaning approach and the constructionist approach is desirable.

In 4.1, I discuss Quine (1960)'s dual conjecture that: a.) count terms have individuation 'built in' and mass terms do not, and b.) this distinction guides ontological learning in children with respect to substance/object distinction. I provide an overview of the literature that refutes the claim that construed ontological commitments are acquired via language acquisition, but rather the substance/object distinction is a pre-linguistic knowledge system that forms part of the broader faculty of language. Quine's second conjecture, that count terms have individuation built in, is discussed in the context of an introduction to the lexicalist and constructionist approaches to nominal number.

In §4.2 I focus on the lexico-meaning approach to nominal number, with a particular focus on Grimm (2012b, 2018)'s account of collective/singulative systems. During the discussion, I submit that Grimm is undoubtedly correct in that construed/perceptual realities necessarily influences the collective/singulative class, specifically the judgement that individuals are perceived as coming in groups, herds, swarms *etc.* is prime. I also discuss that even though construals of the world seem linguistically relevant, lexico-meaning approaches to nominal number struggle to satisfactorily capture crosslinguistic variation of number terms due to their reliance on arbitrariness and/or cultural variation.

In §4.3 the focus is on the constructionist approach to nominal number, with a particular spotlight on Mathieu (2012a, 2013, 2014) and Dali and Mathieu (2021b). The discussion concludes that the constructionist approach offers powerful tools which equate the collective/singulative class to the object mass class in number marking languages and general number in generalised classifier languages, all while equating singulative marking to classification. Despite this, there is little room to incorporate the construed reality influence on nominal number coding.

Finally, in §4.4, I lay out the minimal requirements for a theory of nominal number and suggest a synthesis of the strengths of the lexico-meaning and constructionist approaches to nominal number is warranted.

4.1 Setting the scene

4.1.1 Quine's conjectures

There are at least three levels of ontology. The first is the `FUNDAMENTAL ONTOLOGICAL` level, in which the structure of individuals does not concern human evaluation. At the fundamental ontological level, a bikini is a sum of two arbitrary individuals. The second level is the `CONSTRUED ONTOLOGICAL` level, where human perception, construals and evaluation is relevant to the structure of individuals. At the construed ontological level, a bikini counts as a single individual in its pre-theoretical sense, despite its fundamental mereotopological make up as a scattered whole of disjoint parts.[1] The final level is the `LINGUISTIC ONTOLOGICAL` level, which deals with how individuals are represented in natural language. At the linguistic ontological level, the predicate BIKINI is count in English, but Mandarin Chinese BǏJĪNÍ (*'bikini'*) is number neutral uncountable. Similarly, linguistic representation of MSSC individuals cuts across grammatical classes: DUCK in English denotes a set of individuals that denote individual ducks, but Welsh collective HWYAID (duck.COL = *'duck(s)'*) denotes not only MSSC duck individuals but non-connected sums thereof.

The differing levels of ontology have been associated with different areas of scholarship. Kiss et al. (2021, p.8) point out that the fundamental ontological level is of interest to the philosopher, the construed ontological level is of particular interest to the psychologist, and the linguistic ontological level is the domain of interest for the formal semanticist. While these ontologies can be studied separately, they nonetheless interact. The nature of this interaction is long debated. It is generally accepted that there is a disconnect between the construed ontological level of modelling and the linguistic ontological level of modelling with respect to object mass terms, where object mass terms are not linguistically individuated for counting purposes though they are perceptually individuated. Such matters are discussed explicitly by *e.g.* Chierchia (2021) and Rothstein (2010), where there

[1] See §3.4 for an overview of the philosophical grounding for mereotopology and relevant definitions referred to throughout this chapter.

is a distinction between `COGNITIVELY COUNT` (Chierchia) and `NATURAL ATOMICITY` (Rothstein), which are concepts at the construed ontological level, and their grammatical representation at the linguistic ontological level, *i.e.* the count/mass status.

In the history of modelling the connection(s) between levels of ontology, the most influential work of the twentieth century in this regard is that of Willard Van Orman Quine, whose discussions regarding count, mass, and learnability of predicates which have variously shaped the landscape of contemporary linguistic study (see *e.g.* Strawson (1959); Quine (1960, 1969), but also Moltmann (2021); Pelletier (2011) for a modern discussion). Focusing primarily on Quine (1960), there are two strong claims. The first regards the architecture of grammar, *i.e.* the structure of count and mass terms. Quine (1960, §19) posits that count terms have a built-in division of reference: count terms refer to singularities while mass terms do not have built-in division, and so do not refer to singularities.[2] For Quine, the built-in division of reference for nominal terms is purely linguistic, and not a matter of ontological reality, fundamental or construed. As such, Quine recognises the existence of number neutral uncountable terms in English (*e.g.* FURNITURE), and treats them on par grammatically with stuff mass terms.

Quine's second claim concerns ontological learning and the interaction between construed ontologies and linguistic ontologies. According to Quine only once a child has mastered the (syntactic) use of count terms can they be said to have mastered the object and non-object distinction at the construed ontological level. Specifically, Quine argues that for the child learner, the interaction with the world is of unindividuated 'stuff'. As such, any of the child's interactions with water is on par with the child's interactions with its mother; the child only knows interactions of portions of 'mama' or portions of 'water' and the child has no concept that 'water' and 'mama' may be ontologically distinct types of entities. This lack of division of reference results in all terms initially being equated with the grammatical mass category. Over time, the child is said to slowly master the grammatical count/mass

[2]Quine refers to what I call `COUNT TERMS` as `GENERAL TERMS`.

distinction, and through this mastery the child learns divided reference for both grammar and perception. In other words, the count status of 'mama' at the linguistic ontological level guides the child into a reanalysis of mama-stuff into mama-objects at the construed ontological level. With this, Quine takes a stance within the spirit of ontological relativism, which meshed very well with the then-contemporary Sapir-Whorf claim that language influences the perception of reality (see *e.g.* Sapir, 1929; Whorf, 1940a,b, 1956).

> Once the child has mastered the divided reference of general *[count]* terms, he has mastered the scheme of enduring and recurring physical objects.
> (Quine, 1960, p.86)[3]

(1) `QUINE'S CONJECTURES`
  a. Count terms come with division 'built in', mass terms do not.
  b. Children learn the substance/object distinction through the acquisition of the count/mass distinction.

Though Quine's conjectures were made over sixty years ago, the very strong claims have either directly inspired or are the root of innumerable advancements into the philosophy of language, child development, and formal linguistic theory. Quine's position that children learn construed ontological distinctions through the acquisition of language has now come to be generally rejected in the cognitive and developmental literature, at least in its strongest interpretation. The substance/object distinction is now typically thought to form part of a separate pre/extra linguistic cognitive system, though it clearly interacts with language (Hauser et al., 2002; Spelke and Kinzler, 2007).

In modern linguistic enquiry, the extent to which ontologies interact is hotly debated. In this manner, Quine (1960)'s claim that mass terms come with division

[3] Square brackets and italics my own for disambiguation.

built in and count terms do not is either a fundamental feature of formal systems or is rejected in its entirety. What is at stake here is the nature of what I loosely term the `GENERATIVE COMPONENT`, which I take to be the mechanism that determines grammatical count or mass status. With respect to this generative component, the question is: is count reference 'built into' nominal terms a la Quine, or not?

For linguistics who follow a `LEXICALIST` approach to nominal number, Quine is ultimately correct. Lexicalist accounts posit that the difference between count and mass terms is lexically given: division of reference is 'built in' to count, but not mass, terms. Therefore, in the lexicalist approach to nominal number, the generative component is pre-syntactic in nature. Though concrete proposals vary, advocates of lexicalism include *inter alia* Bloomfield (1984); Chierchia (1998a,b, 2010b, 2017, 2021); Grimm (2012b,a, 2018); Khrizman et al. (2015); Landman (2011, 2016, 2020, 2021) and Rothstein (2010, 2011, 2017, 2021). Within the lexicalist approaches to nominal number, many follow a `LEXICO-MEANING` approach, where there is active debate concerning exactly the manner in which construed ontologies influence linguistic ontologies at the level of pre-syntactic nominal number. For example, discussions pertaining to why all substance terms are grammatically mass (Chierchia, 2010b; Landman, 2020), or why collective terms denote individuals that typically come in groups (Grimm, 2012b) are of central importance in the lexico-meaning approach.

On the other hand, the `CONSTRUCTIONIST` approach to number (implicitly) denies Quine's conjecture. For constructionist accounts, all meaning, including countability features, is compositionally derived regardless of construed ontological fact. In other words, the generative component component is located in the syntax; there is no built-in method of individuation, and therefore no inherent link between different levels of ontology that forces any countability features. Typically for the constructionist approach to number, count/mass flexibility is paramount: there are no mass terms nor count terms, only mass uses and count uses of some root concept. In this way, construed ontological distinctions are relevant for nominal number in so much as they regulate the count/mass usage in context. Followers of this approach include *inter alia* Bale and Barner (2009); Borer (2005a); Borer and Ouwayda (2021); Dali (2020); Mathieu (2012a); Martí (2020) and Pelletier (2012). The main differences between lexico-meaning and constructionist approaches are summarised in (2) and (3).

(2) LEXICO-MEANING ACCOUNTS

a. Number features are lexically given. A nominal term is 'default' coded as grammatically count or grammatically mass.

b. There is an inherent connection between construed ontologies and linguistic ontologies that force grammatical coding.

c. The generative component that determines count/mass status applies pre-syntactically.

(3) CONSTRUCTIONIST ACCOUNTS

a. Number features are derived through computational means.

b. There is no inherent connection between construed ontologies and linguistic ontologies that forces grammatical coding.

c. The generative component that determines count/mass status is syntactic.

Before entering this discussion, there is a terminology issue to clarify. This debate is framed in terms of lexico-meaning versus constructionist approaches to number distinctions. There are adjacent debates that are much older than what I present here. Joosten (2003) discusses grammatical classification under the rubric of GRAMMATICAL/ONTOLOGICAL/CONCEPTUAL/CONTEXTUAL approaches to number. In a similar vein, Grimm (2012b) describes the NOTIONAL/MORPHOSYNTACTIC/CONTEXTUAL approaches to nominal number. While there is crossover between Joosten's grammatical view, Grimm's morphosyntactic category, and what I call the constructionist approach to number, they are not necessarily reducible terms. For example, Bloomfield (1984); Borer (2005a); Chierchia (1998b) and Rothstein (2010) take the count/mass distinction to be grammatical. Yet, Borer (2005a) is the seminal constructionist work, while Chierchia (1998b) and Rothstein (2010) are as lexicalist as they come. The debate I will present will therefore not be in terms of whether there is a construed ontological element to the count/mass distinction, as there clearly is some connection. The debate will concern a.) whether the generative component for the count/mass distinction is pre-syntactic and b.) whether construed ontological commitments influence this generative component.

4.1.2 Ontological distinctions are pre/extra linguistic

Soja et al.'s 1991 study is one of the most notable studies that directly challenges Quine's claim that children learn the substance/object distinction through the acquisition of language. Soja et al. proposed a counter claim to Quine's conjecture, positing that children have a pre-existing set of ontological commitments prior to word learning. To test their counter claim, Soja et al. designed a word-learning task for linguistically naïve children (two-year olds) to see if they extended novel terms for novel items to items of the same shape or material. The hypothesis was that if the children extended novel terms on the basis of shape, then they must have pre/extra-linguistic ontological commitments regarding object and non-object status.

In the object trials, children were handed unfamiliar objects (*e.g.* a honey dipper) which was given a novel name using count/mass neutral syntax *e.g.* 'THIS IS MY BLICKET'. Following this, the child was presented with two new novel objects. The first object was similar in shape and size to the original object (*i.e.* the honey dipper), but it was made out of a different material. The second object was similar in material composition to the original object (*i.e.* the honey dipper) but was different in shape. The experimenter then asked, with reference to the new objects, 'POINT TO THE BLICKET'. A similar procedure was done for novel substances trials where the child was shown a novel substance. In the substance trials, the child was presented with two new substances. The first new substance was materially identical to the original substance, but in a different configuration. The second new substance was materially different from the original substance, but in same configuration.

The results of the experiment showed that children extended the novel term in the object trials to objects of the same shape, but in the substance trials they were not consistent in extending the novel term to substances of the same shape. From these results (and a further two experiments) Soja et al. concluded that contra to Quine's conjecture, pre-syntactic infants have a notion of object and shape and use such a distinction to categorise the projection of word meanings.

Since Soja et al.'s work, there have been further cognitive developmental studies that have come to refute Quine's claims in some way or another. In Xu et al. (1999), it was shown that 10-month old infants use spatiotemporal information to individuate objects prior to learning the kind information/terms for said objects. Such evidence is contra to a Quineian view, as children should not be able to individuate without knowing the term for that object. In another experiment by Carey and Xu (2001), infants are shown to have numeral cognition. Again, this is contra to Quine's claims, as numeral cognition requires individuation of definable objects.

Not only does cognitive development literature refute Quine's conjecture that children are born as blank states with no construed ontological commitments to objecthood, but so does literature concerning studies on adults. Soja et al. (1991); Imai and Gentner (1997) and Barner and Snedeker (2005) argue that if Quine's conjecture is correct, then any Quineian effects should persist into adulthood on words with known meanings: number neutral uncountable terms should not allow reference to objects, as these terms should have no division of reference built-in.[4] In the extreme case, this would predict that languages with no grammatical count/mass distinction (such as Mandarin Chinese) cannot refer to objects at all. Further, due to the conflation of Quine's conjectures, speakers of languages like Mandarin Chinese should not be able to make a distinction between substances and objects at the construed ontological level, as Mandarin Chinese has no grammatical structures to guide learning of these categories. These are very extreme predictions which have been thoroughly tested.

[4]In Quine (1960)'s account, the number neutral uncountable category is an archaic category left over after the acquisition of count terms.

The Quineian prediction that there is a lack of object reference for number neutral uncountable terms was tested explicitly by Barner and Snedeker (2005). In their experiment, Barner and Snedeker (2005) investigated whether objects were accessible for English number neutral uncountable (*i.e.* object mass) terms in count comparison constructions (see §3.5). In their first experiment, Barner and Snedeker presented adults with a photo of two characters. In the object trials, character α was holding a single large object (*e.g.* a fork or shoe), and character β held three smaller objects of the same kind. In substance trials, the set-up was similar: character α had an amount of a substance (*e.g.* toothpaste), and character β had three portions of that substance. In both object and substance trials, the three portions of the substance/objects had a combined smaller volume than the large singular object/substance. Participants were asked, in both trials, the count/mass neutral phrase 'Who has more X?'. The value of X was the experimental condition, which varied between a number neutral uncountable term *e.g.* silverware (fork condition), a count term *e.g.* shoes (shoe condition) or a substance mass term *e.g.* toothpaste (toothpaste condition). The results of the experiments showed that participants based quantity judgements for count and number neutral uncountable terms on the number of individuals significantly more than they did for substance mass terms. Barner and Snedeker concluded that these results indicate that, contra to Quine's claims, some mass terms (specifically object mass terms) have standards of individuation built into them, as the objects were linguistically accessible, and these objects are relevant in (at least) numerosity judgements.

The results of Barner and Snedeker (2005)'s experiments suggest a rejection of the strong Quineian approach to the intersection of the construed ontologies and linguistic ontologies. Nevertheless, there is work to support a weaker Quineian approach. In Lucy (1992) (see also Lucy and Gaskins (2001, 2003) for replications), speakers of English and speakers of Yucatec Mayan (a classifier language) were presented with a similarity judgement task. When showed a cardboard box and asked whether it matched an entity that was similar in shape (*e.g.* a plastic box) or material (*e.g.* a sheet of cardboard), English speakers equated the cardboard box with the plastic box, and Yucatec speakers equated the cardboard box with the material cardboard. Lucy's interpretation of these results is that English speakers consider object denotations as primary while Yucatec speakers considered the material substance as primary. If it is the case that there is a Quineian effect where linguistic ontologies shape construed ontologies, then this result is expected: Yucatec Mayan nominal terms have no individuation (hence classifiers), and this influences percepts/biases of the speakers of that language. Lucy's experiment is not the only one to find such results. Similar observations are found in Imai and Gentner (1997)'s modified replication of Soja et al. (1991)'s experiment, where not only English, but also Japanese speaking infants were tested on a word extension task. The results showed that Japanese speaking infants were more likely than English speaking infants to interpret a novel word as referring to a novel item's substance rather than the the object as an integrated whole. As Japanese is a generalised classifier language with no grammatically count nouns, this result may be expected under weak Quineism.

Are Lucy (1992)'s and Imai and Gentner (1997)'s experiments evidence that speakers of classifier languages may perceive/bias ontological distinctions differently than speakers of number marking languages? Such a conclusion is not clear-cut. Barner et al. (2009, 2010) argue that Imai and Gentner's results may be due to effects of `THE LEXICAL STATISTICS HYPOTHESIS` (Gleitman and Papafragou, 2005; Samuelson and Smith, 1999, *a.o.*). The lexical statistics hypothesis predicts that speakers of a language have knowledge of how terms typically refer in their language, and use that knowledge to make implicit judgements. For example, English speakers implicitly know that a.) English has more count terms than mass terms, and b.) count terms typically refer to objects. Using this knowledge, speakers of English make an implicit judgement that a novel word is grammatically count, and this in turns guides deduction that the novel word refers to an object. Barner et al. (2009, 2010) argue that such a process is not induced in Japanese; there is no count/mass distinction in Japanese and so there is no guiding knowledge for the deduction. Therefore, Japanese speaker judgements are predicted to show no significant difference between assignment of object or substance status to novel terms. If the lexical statistics hypothesis is correct, then the word-extension effects seen in Lucy (1992) and Imai and Gentner (1997) may not in fact due to construed ontological differences between speakers, but due to speakers using knowledge of their language to guide deduction.

Experimental support for the view that speakers of different languages do not differ on perceptual construals is apparent. It is repeatedly shown that speakers of number marking languages and classifier languages do not differ on their substance/object construals. In Li et al. (2009), it was shown that in language-free contexts, speakers of Mandarin, Japanese and English do not differ on their perceptions of substances or objects. Similar conclusions were drawn by Barner et al. (2009) who showed in a substance/object judgement task that monolingual speakers of English and monolingual speakers of Japanese do not differ on their judgements of entities as objects or substances.

With this plethora of evidence, Quine's conjecture that the construed ontological system is developed via the linguistic ontological system does not hold. Indeed, Quine's conjecture faces even more problems when considering that the substance/object distinction is not even inherently human, as primates (Hauser and Carey, 2003; Hauser and Spaulding, 2006; Santos, 2004) and newly hatched chickens (Lea et al., 1996; Valenza et al., 2006) have been shown to have systems of object representation. With all the available evidence, a sound conclusion is as Spelke and Kinzler (2007) summarise: construed individuation is a pre/extra-linguistic trait that forms part of the 'core knowledge systems' in humans and non-humans. It is not a system that is learned through (the acquisition of) language, but rather a system that is biologically endowed, yet is nonetheless related to the broader faculty of language in the Hauser et al. (2002) sense.

4.1.3 Mereological methods for countability

Cognitive and developmental literature shows that, contra Quine (1960)'s claims, the substance/object distinction is a pre/extra-linguistic notion and not learned through the acquisition of language. Modern approaches to the syntax/semantics interface of nominal terms often take these facts as given, and do not agree with Quine's view of grammatical division aiding in the learning of any construed ontological distinction. While this is the case, it does not follow that linguistic ontologies and construed ontologies do not interact. It is clear that the substance/object distinction is active in nominal structures due to the fact that objects are presupposed for the use of `STUBS` and comparison constructions.

Though construed ontologies and linguistic ontologies are somehow intertwined, the manner in which the systems are intertwined is still under debate. Since Godehard Link's seminal 1983 paper, there is a venerable tradition of using mereological methods in nominal semantics to capture this supposed connection, and to also use this connection to predict the grammatical behaviours of count, mass, and plural terms. In Link's original proposal, the universe of discourse, E, has the structure of a Boolean semi-lattice.[5] For demonstration purposes, let's assume a toy universe where there exists only three ducks, a = Alice, b = Bella and c = Carlisle. In this toy universe, the domain of discourse is a mereology that contains these ducks and also their sums.

Under Linkian tradition, the denotation of a one-place predicate, P, is a subset of E which comprises the atomic elements of E. With respect to our toy universe, the predicate DUCK denotes only the individuals of the universe of discourse that are ducks, *i.e.* the set in (6a), which is also visually represented in the bottom tier of the mereology in Figure 4.1. The greatest innovation of Link's proposal is the plural operator, *, defined in (4). In prose, * is a function of algebraic closure upon singularities, creating a join-subsemilattice of E which includes not only the original singularities but also their sums, as in (6b), which is represented by *P in

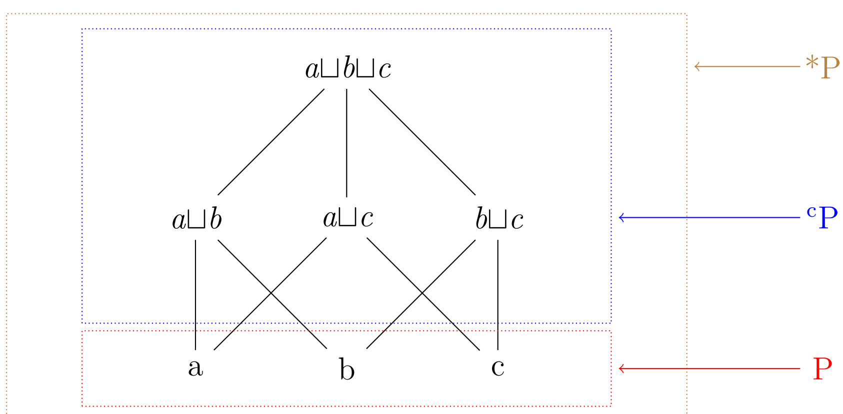
[5] See §3.3 for a background on Boolean Algebras.

the mereology in Figure 4.1. Note, while * is the function of algebraic closure, the notion of plurality in the relevant linguistic sense is not necessarily *. Rather, there is a notion of `PROPER PLURAL`, ^{c}P, as defined in (5). ^{c}P gives the non-atomic (join only) elements of *P, as shown in (6c), and as ^{c}P in the mereology in Figure 4.1.

(4) `ALGEBRAIC CLOSURE`
$*\mathrm{P} \stackrel{\text{def}}{=} \{\mathrm{x}\colon \exists \mathrm{P'}\ \mathrm{P'} \neq \varnothing \wedge \mathrm{P'} \subseteq \mathrm{P} \wedge \mathrm{x} = \sqcup \mathrm{P'}\}$
The set that contains all sums of things in P
From Champollion and Krifka (2016), adapted

(5) `PROPER PLURAL`
$^{c}\mathrm{P} \stackrel{\text{def}}{=} *\mathrm{P} - \mathrm{P}$
The set that contains all sums of things in P minus the atomic set

(6) a. $\textsc{duck} = \{a, b, c\}$
b. $^{*}\textsc{duck} = \{a, b, c,\ a{\sqcup}b,\ a{\sqcup}c,\ b{\sqcup}c,\ a{\sqcup}b{\sqcup}c\}$
c. $^{c}\textsc{duck} = \{a{\sqcup}b,\ a{\sqcup}c,\ b{\sqcup}c,\ a{\sqcup}b{\sqcup}c\}$



A = Singularities, *A = algebraic closure on A, ^{c}A = proper plural of A

Figure 4.1: Representation of Link's subdomains of discourse

For Link (1983), singular count and plural terms take their reference from the domain of individuals. Substances are said to take reference from the domain D, the domain of portions of matter, which is separate to the count domain. D is said to have Boolean structure, though it is not required to be atomic. Additionally, any individuals relevant to a predicate that are in D are identical. It is within these two domains that the count/mass distinction is said to arise; objects have their extensions in E and are countable, while substances have their extensions in D, and are uncountable. This type of analysis captures that both plurals and mass terms are cumulative (7) as both mass and plural terms denote sums/joins. According to this definition, if two individuals are in the extension of predicate P, then their sum is also in the extension of the predicate P.

(7) CUMULATIVITY:
P is cumulative iff: $\forall x \forall y[x \in P \wedge y \in P \rightarrow x \sqcup y \in P]$
P is a cumulative predicate if when x and y are in P, then the sum of x and y is also in P
From Rothstein (2010, p.10)

Though D and E are separate domains, they are connected via a homomorphism which states that D is a subset of the atoms of E. Therefore, the matter which makes up an individual is represented. A concrete example would be one of the ducks in our toy universe, Alice. While Alice is an atom of domain E, the matter which comprises Alice is represented in domain D. The homomorphism is a mapping function, h, which maps the entities of E to their constituent parts in D. The justification for such a conjecture is that a language may use different nominal expressions (of differing grammatical category) to refer to entities that occupy the same spatiotemporal space. This analysis allows (8) to not be contradictory, as the ring refers to an entity in E, but the stuff making up the ring is the value h, the mapping from E to D. It is by this measure that the inaugural formal mereological approach to nominal number is Quineian in nature as the reference of mass and count terms is 'built into' nominal terms via reference to different domains.

(8) The gold making up Terry's ring is old, but the ring itself is new (not old)

The Linkian approach verbatim did not last long. As pointed out by Bach (1986a), the dual domain system runs into problems with sentences such as (9). In (9), both the snow and the H_2O are constituents of the snowman, and therefore both constitute the mapping function h. According the Link (1983)'s analysis, (9) must be contradictory, though common wisdom indicates that it is not.

(9) The snow making up this snowman is quite new but the H_2O making it up is very old (and the H and the O even older!)
From Bach (1986a, p.13)

Additionally, Chierchia (1998a) and Rothstein (2010) point out issues with the dual domain approach when considering the crosslinguistic distributions of number neutral uncountable terms. First, Chierchia notes that the dual domain approach is problematic from a crosslinguistic perspective as there is variation of count and mass usage for what is arguably the same concept. For example, in English HAIR is a mass term, but the Italian counterpart CAPELLO is a count term. This would necessitate that the same object belongs to two different domains with no mapping function between the uses. Therefore, hair must somehow be simultaneously atomless (English) and atomic (Italian), a questionable mereological feat. In a similar vein, Rothstein (2010) notes there is a paradox when two mass terms, which would have their denotations in the same domain D, refer to the same constituent 'stuff', presumably without any sort of mapping relation h. To see this, in a structure parallel to (9), the use of two mass terms in (10) does not lead to contradiction. Similarly, Doetjes (2021, p.85) points out that wooden furniture can be made of old wood, resulting in a similar paradox. The issues that Chierchia, Doetjes and Rothstein highlight arise precisely because Link (1983)'s model equates grammatical category with construed ontological domains, and as such struggles to capture the fact that number neutral uncountable terms are conceptually individuated though not necessarily linguistically individuated, at least in some constructions.

(10) a. This jewellery is new but the gold it is made of is old
b. The curtaining is new, but the fabric it is made of is old
From Rothstein (2010, p.350)

It is now not common to assume a dual domain approach to count and mass terms in contemporary linguistic enquiry. Rather, in accounts which follow a mereological approach to number, mass and count terms are said to draw their denotation from the same domain. While this is so, Link (1983)'s contribution to nominal semantics cannot be understated, as it is now common practice to assume mereological Linkian inspired approaches to count and mass terms. That is, closure under sum, *, or its variant, ^{c}P, and the cumulative effects it induces are still very much induced in the analysis of count versus plural/mass terms (though *c.f.* non-mereological approaches, *e.g.* Borer (2005a)).

With the stage set, we are about to turn to modern accounts of nominal number from the point of view of collective/singulative languages. As it turns out, the two most developed analyses of collective/singulative systems fall either side of the lexicalist/constructionist debate. The account in Grimm (2012a,b, 2018) (and later Wągiel and Shlikhutka (2023a)) is lexicalist in nature, as they claim that collective/singulative coding is pre-syntactic and inherently linked to construed commitments of proximity. Conversely, the accounts in Mathieu (2012a, 2013, 2014); Dali and Mathieu (2021a) are constructionist in nature, having been inspired by Borer (2005a)'s exoskeletal approach, and have little to no concern of any construed ontological input into the system.

4.2 Lexicalism

Though the dual domain approach is no longer typically adopted, the spirit of mereological modelling of the domain of discourse is the norm in formal lexicalist approaches to number. For these approaches, Quine (1960) is ultimately correct in that count terms come with individuation built in (they denote some equivalent of the atomic subset of the domain of discourse), and mass terms do not (they denote some equivalent of a sub-mereology of the domain of discourse). The difference between mass and count terms is lexically given, and so the generative component that derives count and mass status is pre-syntactic.

4.2.1 The mapping property

Strictly, a lexicalist account need not refer to construed ontological status of relevant individuals for nominal coding (the generative component may be purely grammatical), but many lexicalist accounts do connect the status of construed ontological commitments to grammar. These are what I call the `LEXICO-MEANING` accounts, where the core theme that binds them together is the agreement that the count/mass distinction cannot be completely arbitrary. The unifying claim that underpins this is that all substance terms in all languages are coded as grammatically mass. The reasoning for this is simple: construed ontological structures of pre-theoretical substances do not have (relevant or accessible) individuals, and therefore neither do the terms for them in natural languages, and as such semantic singularity is not encoded in those terms. According to lexicalist tradition, there is no known language which has substance concepts realised as grammatically count in their uninflected/unclassified form (Chierchia, 1998b, 2010b, 2017, 2021; Rothstein, 2010; Doetjes, 2017; Moltmann, 2020). Such an observation is explicitly stated by Chierchia (2010b) who proposes `THE MAPPING PROPERTY` as a linguistic universal.

> `THE MAPPING PROPERTY`:
> In any language L, substances are coded as mass by the tests prevailing in L. The idea is that each language will have specific morphosyntactic generalizations that distinguish mass from count (just like, say, every language has criteria to tease subjects from objects). By those tests whatever they turn out to be, in no language the basic words for say blood or air will come out as count. This is an extremely strong and substantive universal.
> (Chierchia, 2010b, p.5)

That all substances are coded as mass terms in all languages is a strong claim. It is also a claim that is yet to be convincingly refuted under lexicalist assumptions. The closest evidence there is to a denial of the mapping property comes from Lima (2010, 2014)'s work on Yudja, a language which allows numerals to freely combine with any term, including substance terms. For example, in the case that numerals combine with substance denoting terms, two readings may be obtained: a packaging reading, and a reading of 'unconventional contexts'. The unconventional reading is shown in (11), where the reading distributes over three blood-dripping events.

(11) Context: João cut his finger and three drops of blood fell on the floor: one near the river, one near the house and another near the school.

Txabïu apeta pe∼pe∼pe — Yudja
three blood drip-RED
'*Three (drops of) blood dripped*'
From Lima (2014, p540-541)

As the valid combination of terms with numerals (with no intervening structure) is perhaps the only universal property of grammatically countable terms (see Chierchia, 2010b), such evidence points to Yudja as being a a language with grammatically count substance terms. However, the case for Yudja being an all-count language is not clear. Chierchia (2015, 2021) argues that countable mass in Yudja is illusionary and not evidence for a contradiction of the mapping property. Part of Chierchia's argument comes from the behaviours of Yudja quantifiers, which are interpreted differently with respect to object and substance denoting terms. For example, when appearing with a substance term URAHU is interpreted as a quantifier of amount (12), but when appearing with an object term, URAHU is interpreted in terms of size. From this, Chierchia argues that if Yudja has no count/mass distinction, then the quantifiers should be interpreted the same in each contexts. Yet, they cannot be. This indicates at the very least that substance terms do not have commitments to a structure which contains pre-existing individuals, and if there are no commitments to pre-existing individuals, then the term cannot be count. As such, Chierchia argues that the correct analysis of Yudja is to treat the language as indeed having a count/mass distinction, and such uses of 'countable mass' are actually instances of covert classifiers. With this, it looks like Yudja may not in fact be a counterexample to the mapping property (though *c.f.* Doetjes, 2021).

(12) Urahu ahuanama txa — Yudja
A lot milk
'*A lot of milk in a single place*'
From Lima (2010, p.160)

(13) Urahu ali — Yudja
big child
'*The child is big*'
From Lima (2010, p.160)

Assuming the mapping property holds, there are clear and strong predictions regarding the interaction between construed ontologies and linguistic ontologies. Specifically, the lexico-meaning pre-syntactic generative component which determines count/mass status is such that it necessitates that all substances are coded as mass terms. With this, there is a venerable tradition within the lexicalist camp of nominal number to identify further intersections between construed ontologies and linguistic ontologies, *i.e.* further sensitivities of this generative component. Such a task is no mean feat.

First, any lexicalist account must answer why the mapping property is asymmetric: the converse of the mapping property does not hold. It is not the case that object concepts must be coded as grammatically count, as clearly exemplified by the existence of number neutral uncountable terms including object mass in number marking languages, collectives in singulative languages and general number in generalised classifier languages. Though grammatical countness is not a necessary outcome for object properties, the objecthood construal must nonetheless be present in object denoting nominal terms, despite grammatical category, to account for the behaviour of `STUBS` and comparison constructions (*c.f.* Chierchia, 2021, p.36). Therefore, any lexicalist approach must look beyond the substance/object distinction for any satisfactory account of construed ontological/linguistic ontological mapping.

A second challenge for lexicalist approaches to number requires the defining of the mapping property in such a way that allows all object properties to be coded as number neutral uncountable in generalised classifier languages, but allows only some object properties to be coded as number neutral uncountable in number marking languages and singulative languages. Such a task is extraordinarily difficult due to the fact that in number marking languages and singulative languages there is no set of core number neutral uncountable terms. As already discussed, Chierchia (1998a) pointed out that English HAIR is number neutral uncountable, while Italian CAPELLO/I (hair/PL = *'hair/s'*) is count. Finding examples of crosslinguistic mismatches is not difficult. Sutton and Filip (2016c) note that Bulgarian LEŠTA (lentil.NNU = *'lentil(s)'*), BOB (bean.NNU = *'bean(s)'*) and Czech ČOČKA (lentil.NNU = *'lentil(s)'*) are number neutral uncountable terms while English and Finnish equivalents are count terms. There are also languages which have been argued to have no number neutral uncountable terms, such as Greek (Tsoulas, 2009; Chierchia, 2010b, 2021).[6]

The variation of coding of number neutral uncountable terms is advanced by collective/singulative languages. Not only do inventories of collective terms differ from object mass terms in *e.g.* English, but they also differs across collective/singulative languages. For example, while TUFAAḤ (tufaaḥ.COL = *'apple'*) is collective/singulative in (the dialects of) Arabic, the equivalent AFAL (afal.SG = *'apple'*) is singular/plural in Welsh. Reversing the roles, CACWN (wasp.COL = *'wasp(s)'*) is collective/singulative in Welsh while DUBBAR (wasp.SG = *'wasp(s)'*) is singular/plural in (Modern Standard) Arabic.

[6] *c.f.* Alexiadou (2015); Erbach (2021) who claim that Greek has a very small number neutral uncountable inventory.

Further, even within a singulative language, there is variation amongst speakers with regards to whether a given term is in the singular/plural class or collective/singulative class. For example, in Welsh there is variation on use of MALWOD (*'snail(s)'*). In South Wales dialects, MALWOD is a collective term, contrasting against a singulative MALWOD-EN. In North Wales dialects MALWOD is the plural counterpart to a faux singulative MALWEN, where -OD/-EN form a suppletive singular/plural marking pattern.[7] A similar observation is seen in the dialects of Arabic, though in this case the lexical form of the word is different. In modern standard varieties, DUBBAR (*'wasp'*) is part of the singular/(broken) plural class with plural form DABAABIIR. According the The Living Arabic Project[8], in Levantine dialects the equivalent terms are part of the collective/singulative class, with ZURQUṬ (*'wasp(s)'*) being a collective and ZURQUT-AH the singulative.[9]

The picture painted of variation in the coding of number neutral uncountable terms is complex, and is summarised in Table 4.1. With this picture, the challenge for any lexico-meaning account must include some prediction of how and why the mapping property (and its failed converse) manifests if the generative component within the lexicon is indeed sensitive to construed ontologies. Whatever shape a lexicalist account takes, it must somehow exclude all mass and number neutral uncountable terms from combining with numerals, but also allow objects denoted by number neutral uncountable terms to be accessible for the combination with `STUBS` and numerosity interpretations in comparison constructions.

[7] See §2.8 for more details on faux singulatives.

[8] The Living Arabic Project is an online crowd-funded dictionary of the dialects of Arabic with a focus on Egyptian and Levantine dialects. Available online: https://www.livingarabic.com/en

[9] While I have assigned certain forms to certain dialects, this does not entail that all speakers of dialects use these terms.

	Count	**Number neutral uncountable**
English	LENTIL-S WASP-S BEAN-S SNAIL-S APPLE-S CAT-S	CORN
Welsh	AFAL-AU (*'apple'*) MALWEN (*'Snail'*) (FS, NW) CATH-OD (*'cat'*)	YD (*'corn'*) CORBYS (*'lentil'*) CACWN (*'wasp'*) FFA (*'bean'*) MALWOD (*'snail'*) (SW)
Arabic	DABBAR (*'wasp'*) QIṬ (*'cat'*)	DUR (*'corn'*) 3DAS (*'lentil'*) ZUQRUṬ (*'wasp'*) (LEV) FUUL (*'bean'*) TUFAAḤ (*'apple'*)
Czech		ČOČKA (*'lentils'*)
Finnish	LINSSI-T (*'lentils'*) PAVU-T (*'beans'*)	
Bulgarian		LEŠTA (*'lentils'*) BOB (*'beans'*)
M. Chinese		YÙMǏ (*'corn'*) BIǍNDÒU (*'lentils'*) HUÁNGFĒNG (*'wasp'*) DÒU (*'bean'*) PÍNGGUǑ (*'apple'*) WŌNIÚ (*'snail'*) MĀO (*'cat'*)
Russian		FASOL' (*'bean'*)

SW = Southern Welsh dialects, NW = Northern Welsh dialects, FS = faux singulative, LEV = Levantine dialects

Table 4.1: Crosslinguistic variation of the count/number neutral uncountable class

4.2.2 Vagueness and overlap

Over the years, there have been attempts to capture the construed ontological nature of the lexico-meaning generative component for the count/mass distinction. One way in which a lexicalist approach to nominal number could play out is in terms of `SUPERVALUATIONISM`, as in Gennaro Chierchia's most recent 2010b, 2017 and 2021 accounts. The central claim in *e.g.* Chierchia (2010b) is that construed ontological `VAGUENESS` of individuals ultimately underlies the pre-syntactic generative component.

The vagueness argument boils down to stating that the atomic individuals which make up the mereological structure of water and rice are (construed as) `VAGUE`, in that what counts as an atom is not consistent across all contexts. In contrast, a prototypical cat atom is not vague in all contexts, but rather it is stably atomic. Chierchia claims that this inherent construed vagueness of individuals affects the way we count. Properties which do not have stable atoms in their positive extensions in all contexts are mapped to mass terms. With this, it is clear why the mapping property holds: substances will never be stably atomic across contexts, and as the pre-syntactic generative component is sensitive to stable atomicity, then substances will always be coded as grammatically mass.

But what about number neutral uncountable terms under the vagueness account? Despite the similarities between English object mass and Mandarin general number terms, Chierchia takes them to be fundamentally different beasts. Chierchia argues that, clearly, the individuals denoted by FURNITURE are no more vague than the individuals denoted by CAT and so a vagueness approach cannot be appropriate to predict the countability status of object mass terms. To account for object mass predicates in number marking languages that are not vague, Chierchia claims this is simply a `COPY-CAT` effect on canonical stuff mass terms, terming them `FAKE MASS`. For generalised classifier languages like Mandarin Chinese, the object mass effect is illusionary. There are no object mass terms, but all terms are coded as kinds, essentially sidestepping a vagueness-based pre-syntactic generative component all together. Therefore, in Chierchia's account, construed ontological distinctions are predicative of object mass terms only in number marking languages, and even then there are exceptions.

Another way in which a lexicalist approach to nominal number could be stated is in terms of vertical overlap such as the `ICEBERG SEMANTICS` approach pioneered by Fred Landman and associates (Landman, 2011, 2016, 2020, 2021; Khrizman et al., 2015). In Iceberg Semantics, the pre-syntactic generative component is not sensitive to (vague) atomicity. Rather, `DISJOINTEDNESS` is key.

Under the disjointedness view, what counts as an individual in the pre-theoretical sense must be disjoint at every perspective, and it is this disjointedness which leads to count-coding. To see how this works, consider that in any perceivable context, any given cat will be disjoint from any other instance of a cat.[10] For substances such as water, there is no disjoint entity that counts as a pre-theoretical individual at every context. The pre-syntactic generative component is sensitive to this distinction where disjointedness at every context leads to count coding, mass otherwise. Therefore, the predicate CAT will be grammatically count while WATER will be grammatically mass.

To account for artificial collection object mass terms[11] *e.g.* FURNITURE, KITCHENWARE, IRONWARE, POTTERY under disjointedness, Landman invokes `VERTICAL OVERLAP`. To exemplify, Landman discusses the term POTTERY. Heavily paraphrasing Landman (2020), we may assume that the basic entities that make up POTTERY comprise the disjoint set in (14a). Yet, we can also consider the set in (14b) as a valid instance of the entities that comprise POTTERY. Notice that the set in (14b) is vertically overlapping: *the saucer* appears as a minimal element of the set, but also as part-of another element of the set, *the-cup⊔saucer*. As such, *the cup* and *the-cup⊔the-saucer* simultaneously may both count as individuals, *i.e.* as *one*. It is exactly this lack of disjointedness at every context that Landman (2020) argues leads to number neutral uncountable coding.

[10] These contexts are based on normal prototypical use of disjointedness, and as such the case of conjoined kittens is not considered an exception.

[11] `NEAT MASS` in Landman's terms.

(14) a. Pottery Atoms = {the teapot, the cup, the saucer, the fruit bowl}
b. Pottery items = {the teapot, the cup, the saucer, the fruit bowl, the cup and saucer, the teaset}.

The disjointedness approach for number neutral uncountable in *e.g.* Mandarin Chinese, Landman (2020) is similar to Chierchia in that the issue of connecting construed ontological construals to linguistic ontological is completely sidestepped: Mandarin Chinese is said to arbitrarily neutralise the count/mass distinction. Arbitrariness on number neutral uncountable coding is also said to apply to select terms in *e.g.* English where the objects denoted are not clearly overlapping, such as POULTRY.

While lexico-meaning accounts such as Chierchia (2010b) and Landman (2020) are powerful in that they provide an answer to why the mapping property holds as well as the appearance of number neutral uncountable terms, they do not provide a fully satisfactory account of the interface of construed ontologies and linguistic ontologies. A common criticism of lexicalist views is that despite their generative power, they are not fully predicative of crosslinguistic coding of number neutral uncountable terms. It has been pointed out by Landman (2020); Sutton and Filip (2016d); Erbach (2021) that vagueness alone cannot account for the count/mass distinction as it is not immediately obviously why RICE is coded as a number neutral uncountable term in English, but PEA, LENTIL, BEAN are are count. The situation is compounded by the fact that these terms are number neutral uncountable in other languages, as already shown in Table 4.1.

The disjointedness approach to nominal number coding does not fare much better. It is not clear why some languages such as English seemingly have (near) compulsory number neutral uncountable coding for vertical overlap, while others do not. For example, while English FURNITURE is count, French MEUBLE/MUEUBLES (*'furniture/s'*) is count. If it is the case that vertical overlap leads to number neutral uncountable coding, why should MEUBLE be count? Further, in Dutch, as Landman (2020, 2011) points out, there is a three-way distinction between MEUBEL/MEUBELS which are singular/plural and MEUBILAIR which is number neutral uncountable. Perhaps we could posit that some languages are sensitive to vertical overlap (English), others are not (French), and some others allow both methods (Dutch). If this is the case, then it is not clear why terms such as PORTION are count in English. As pointed out by de Vries and Tsoulas (2024, p.411), English PORTION does not clearly denote a vertically disjoint set. Paraphrasing de Vries and Tsoulas's argument, imagine a scenario where we have three bowls of soup and three pieces of buttered bread. How many portions are there? Three, or six? Under Iceberg Semantics, PORTION should not be countable, yet it is, as clearly exemplified by (15).

(15) We ordered three portions for Edward, and two for Bella English

Interestingly, both Chierchia and Landman are seemingly aware that neither vagueness nor disjointedness respectively can account for the whole picture. In a footnote, Chierchia (2010b, p.32) acknowledges that the copy-cat effect of fake mass terms is 'totally arbitrary' and may be driven by a 'lack of interest in the atoms'. Similarly, Landman (2020, p.194) treats number neutral uncountable terms such as POULTRY as arising through lexical choice, though no explanation why this is so is offered.

One solution to the failure of vagueness alone or disjointedness alone would be to synthesise the intuitions such that either vagueness *or* overlap could lead to mass coding. This is exactly the approach taken by Sutton and Filip (2016a,b,c,d) who propose a dual-source analysis of the count/mass distinction, where vagueness and overlap may independently lead to mass coding, and if a term is *both* vague and overlapping (*i.e.* substances) then mass coding is necessitated. Yet, while Sutton and Filip's approach is definitely a step forward in our understanding of count/mass coding, it alone is not sufficient to predict the coding of collective/singulative languages. That is, it is plainly obvious that *ducks* are neither vague nor overlapping, and yet HWYAID (duck.COL = *'duck(s)'*) and BAṬ (duck.COL = *'duck(s)'*) are both number neutral uncountable (collective) terms.

4.2.3 Individuation

If neither (construed) vagueness, overlap, nor these features together can be taken to be the construed ontological source of the pre-syntactic generative component which underlies count/mass distinction in the lexicalist tradition, then what is? One candidate for a meaning based lexicalist approach to number is Wierzbicka (1988)'s claim that `EASE OF DISTINGUISHABILITY` and `MANNER OF INTERACTION` with entities is relevant for nominal coding.

Wierzbicka argues that peas are bigger than grains of rice, and as such PEA is count while RICE is mass in English. With respect to manner of interaction, Wierzbicka proposes that ONION, RADISH, OLIVE are count and GARLIC is mass is because of an apparent tendency to eat radishes, olives and onions (but not garlic) whole. In Russian, however, LUK (*'onion'*) is number neutral uncountable. Wierzbicka suggests that this is due to eating habits as:

> radishes are often eaten individually, whereas onions are often chopped (Wierzbicka, 1988, p.502).

If Wierzbicka is right and manner of interaction is linguistically relevant, this would predict that cross-linguistic differences may arise due to cultural differences; *i.e.* if interaction with a given entity differs substantially across cultures, then the grammars representing those entities will also differ. This is perhaps a fruitful line of enquiry, considering the variation seen inter and cross linguistically in Table 4.1. Though potentially promising, Wierzbicka's claims are not without scrutiny. For one, Palmer (1990) argues against Wierzbicka's account, claiming that it is impossible to know how people and cultures construe interaction with certain entities, and as such using such an argument to explain nominal number leads to circularity. Similarly, Zwicky (2001) rejects postulating the count/mass distinction is based in how individuals are conceptualised as there is no independent evidence for such a conjecture.

4.2.3.1 Experimental evidence for individuation

Despite the criticisms, Wierzbicka's intuitions have been partially vindicated by Middleton et al. (2004)'s series of experiments which aimed to determine whether there is a systematic relation between construals of individuation and count/mass coding (in English). For Middleton et al. individuation is a percept which consists not only of `SIZE` of elements, but also `SPATIAL CONTIGUITY`, *i.e.* how close together the elements of an aggregate are.

In one experiment (their experiment 3), Middleton et al. showed that spatial continuity was a relevant feature for nominal coding, at least in a task where participants labelled novel terms as count or mass. In this experiment, participants read a phrase containing a novel term. The novel term appeared either with count syntax ('These things are called briffs') or mass syntax ('This stuff is called briff'). Participants were then presented with a pair of images depicting novel aggregates which differed with respect to size and/or proximity but were otherwise identical. After being presented with a pair of novel aggregates, participants were asked which image they thought the sentence with a novel term was describing. An example of Middleton et al. (2004, p.383)'s pairs can be seen in Figure 4.2, and the pairs are summarised in (16), were condition labels match the labels in Figure 4.2.

(16) Middleton et al.'s Experimental Conditions:

Proximity condition

Condition C:	Large-Apart	-	Large-Close
Condition E	Small-Apart	-	Small-Close

Size condition

Condition D:	Large Apart	-	Small Apart
Condition B:	Large-Close	-	Small-close

Proximity + Size condition

Condition A:	Large-Apart	-	Small-Close

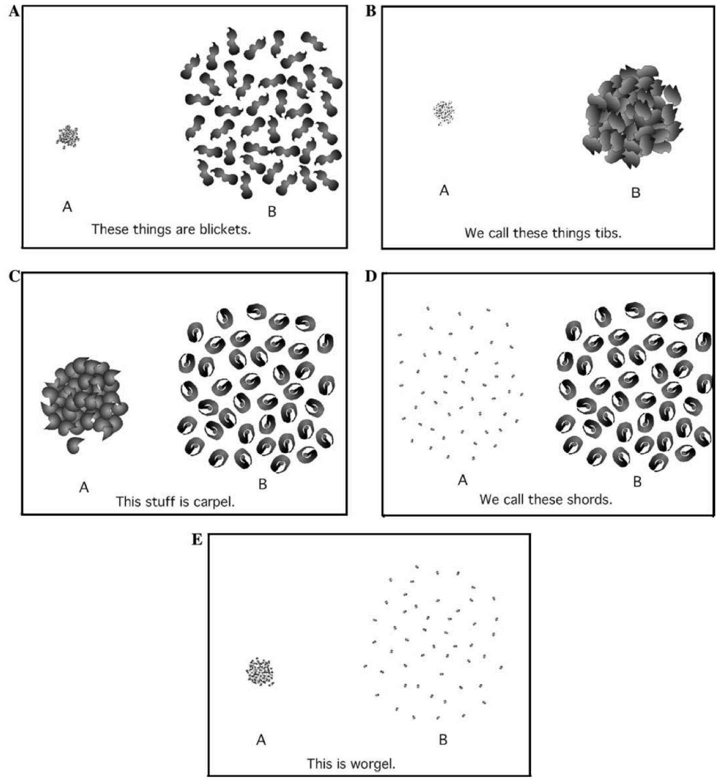


Figure 4.2: Middleton et al.'s stimuli

The results of Middleton et al.'s experiment showed that when only proximity varied (conditions C and E), participants were significantly more likely to assign mass syntax to the aggregate where individuals were close together. The inverse was also found for count syntax choices. In the conditions where only size differed (conditions D and B), participants were not more likely to choose the large-apart and large-close aggregates over small-apart and small-close aggregates, respectively, for neither mass nor count syntax. Finally, in the condition where both size and spatial contiguity differed (experimental condition A), participants were more likely to assign mass syntax to the image with an aggregate that constituted small-close entities than to the image with large-apart entities. Again, the inverse was found for count syntax choices. From these results, Middleton et al. concluded that proximity alone is a significant factor in the online choice of count and mass coding, as is a mix of proximity and size, but size alone is not.

Though size was not significant in participants' choices of count and mass syntax, Middleton et al. tested the same pairs in Figure 4.2 but for distinguishability. Middleton et al. instructed participants to choose the aggregate which:

> contains individual elements that are easier to see, or 'pick out' and distinguish from each other.
> (Middleton et al., 2004, p.386)

In these trials it was shown that size alone (conditions D and B) is a significant factor in participants' choices, where larger element aggregates were chosen significantly more than smaller element aggregates. With this, Middleton et al. conclude their experiment by suggesting participants' coding of novel terms as count or mass are not affected by size even though size itself is a factor in ease of distinguishability. Such results are argued to support Wierzbicka (1988)'s claim that ease of distinguishability (defined as aggregation) is relevant to the count/mass distinction, for if individuals are aggregates, they are more likely to be be coded as mass terms. At the same time, Wierzbicka's claim that size is predicative of the count/mass system was not supported.

Middleton et al. investigated not only Wierzbicka's claim that ease of distinguishability is predicative of the count/mass coding, but also Wierzbicka's claim that the manner of interaction is, too. In Middleton et al.'s fourth experiment, participants were split into two groups: baseline participants and interaction participants. All participants were presented with a box which contained a novel aggregate (yellow sand-like material). The baseline participants were told to choose a phrase with mass or count syntax that may be used to describe the contents of the box. The interaction participants were informed to watch an experimenter silently pick up individual grains of the aggregate. The individuation participants were told to copy this task for five minutes. At the end of the task, the individuation participants were also asked to choose between phrases with count or mass syntax to describe the material they interacted with. Middleton et al. found that the interaction participants chose novel count terms to describe the aggregate significantly more than the baseline participants. Middleton et al. take this result to indicate that interaction with the individuals of an aggregate increases the likelihood of count syntax choices, supporting Wierzbicka's claim that interaction plays a role in count/mass nominal coding.

With the results of Middleton et al.'s experiments, it seems that Wierzbicka's claims have some credence, despite Palmer (1990)'s criticisms. Specifically, individuation modelled as aggregation and interaction (but not necessarily size) seems to be relevant to the the online coding of novel terms as count or mass.

4.2.3.2 Individuation and collective/singulative systems

If individuation (defined as aggregation and manner interaction) underlies the lexico-meaning pre-syntactic generative component, number neutral uncountable terms in languages which do not have obligatory mass-coding (*i.e.* number marking languages and singulative languages) should denote aggregate concepts. This observation is borne out for natural kind object mass terms in English, which typically denote cereals, grains, and pulses (*e.g.* BARLEY, RICE, MILLET). As it turns out, in the collective/singulative literature it is generally accepted that terms in the collective/singulative class typically refer not only to granular aggregates, but also individuals that come in groups, collections, swarms, herds, piles *etc.* (Asmus and Werner, 2015; Grimm, 2012b, 2018; Haspelmath and Karjus, 2017; Hemon, 1975; Jaradat, 2023a; Jaradat and Jarrah, 2022; Nurmio, 2017; Roberts and Gathercole, 2012; Stolz, 2001; Wągiel and Shlikhutka, 2023a, *a.o.*).[12]

The non-exhaustive examples in Table 4.2 and Table 4.3 show that the uniting feature of collective terms in Welsh and Hejazi/Modern Standard Arabic is indeed a notion of construed aggregation.[13] There is a more exhaustive list of collective/singulative terms appendix A.

[12] The same observation is described as 'gregarious in number' by Haspelmath (2008).

[13] Similar observations are made by Dimmendaal (2000) for the Turkana singulative class, Unseth (1988) for the Majang singulative system, by Grimm (2018) for the Dagaare singulative class, and by Wierzbicka (1988) for Russian. By collating different sources, Haspelmath and Karjus (2017) also make the same point about Surmic, Krongo, Cushitic and Kadugli.

Seeds / Cereals			Insects		
Col	**Sing**	**Gloss**	**Col**	**Sing**	**Gloss**
ceirch	ceirich-yn	oat(s) → oat	cacwn	cacyn-en	wasp(s) → wasp
dincod	dincod-yn	seed(s) → seed	mogrug	mogrug-yn	ant(s) → ant
pys	pys-en	pea(s) → pea	egnod	egnod-yn	flea(s) → flea
cnau	cnau-en	nut(s) → nut	lleu	lleu-en	lice → louse
Vegetables / Fruits			**Flora**		
Col	**Sing**	**Gloss**	**Col**	**Sing**	**Gloss**
afan	afan-en	raspberr(ies) → raspberry	coed	coed-en	tree(s) → tree
gellyg	gellyg-en	pear(s) → pear	ysgaw	ysgaw-en	elder(s) → elder
moron	moron-en	carrot(s) → carrot	helyg	helyg-en	willow(s) → willow
madarch	madarch-en	mushroom(s) → mushroom	gellesg	gellesg-en	iris(es) → iris
Small/medium sized animals			**Misc**		
Col	**Sing**	**Gloss**	**Col**	**Sing**	**Gloss**
llygod	llygod-yn	mouse → mice	sêr	ser-en	star(s) → star
moch	moch-yn	pig/s → pig	plant	plent-yn	child(ren) → child
pysgod	pysgod-yn	fish(es) → a fish			
hwyaid	hwyad-en	duck(s) → duck			

Table 4.2: Welsh collective/singulative terms denote aggregate concepts

Seeds / Cereals			Insects		
Col	**Sing**	**Gloss**	**Col**	**Sing**	**Gloss**
lawz	lawz-ah	almond(s) → almond	naḥl	naḥl-ah	bee(s) → bee
jawz	jawz-ah	walnut(s) → walnut	naml	naml-ah	ant(s) → ant
fuul	fuul-ah	bean(s) → bean	qaml	qaml-ah	lice → louse
3das	3das-ah	lentil(s) → lentil	dhubaab	dhubaab-ah	flie(s) → fly
Vegetables / Fruits			**Flora**		
Col	**Sing**	**Gloss**	**Col**	**Sing**	**Gloss**
tufaaḥ	tufaaḥ-ah	apple(s) → apple	shajar-ah	shajar	tree(s) → tree
'ijaaṣ	'ijaaṣ-ah	pear(s) → pear	Safaaf	Safaaf-ah	willow(s) → willow
shamandar	shamandar-ah	beetroot(s) → beetroot	naxal	naxl-ah	palm(s) → palm
fuṭr	fuṭr-ah	mushroom(s) → mushroom	zahr	zahr-ah	flower(s) → flower
Small/medium sized animals			**Misc**		
Sing	**Col**	**Gloss**	**Sing**	**Col**	**Gloss**
dajaaj	dajaaj-ah	chicken(s) → chicken	beyd	beyd-ah	egg(s) → egg
samak	samak-ah	fish(es) → fish			
baṭ	baṭ-ah	duck(s) → duck			
baqar	baqar-ah	cow(s) → cow			

Table 4.3: Arabic collective/singulative terms denote aggregate concepts

If indeed the lexicalist generative component is sensitive to construed aggregation, some caveats must be made. First, any construed aggregation sensitivity of the pre-syntactic generative component can only apply in some, but not all languages. No matter how precise the theory of construed aggregation is for the number marking and singulative languages, it is completely irrelevant for generalised classifier languages as all terms are coded number neutral uncountable. Relating to this, there must be an account as to why different languages are sensitive to different levels of construed aggregation. In singulative languages like Welsh and Hejazi/Modern Standard Arabic, number neutral uncountable (collective) coding for fruits, vegetables, insects and small animals is highly productive, but in number marking languages such as English number neutral uncountable (object mass) coding applies only to granular aggregates with any efficiency.

Second, there needs to be an account for the sheer number of exceptions to a construed aggregation account of the count/mass distinction. It has long-time been established that similar individuals have varied count/mass coding, for example OATS/WHEAT, BEANS/RICE and VEGETABLES/FRUIT where the first item in each pair is count, and the latter is object mass (Bloomfield, 1984; Wierzbicka, 1988; Ware, 1979; Gleason, 1969). If (granular) aggregation is relevant to the English count/mass distinction, then why are there exceptions to this rule? More examples are shown in Table 4.4.[14]

The situation is paralleled in the collective/singulative class, where Welsh and Hejazi/Modern Standard Arabic shows that one singulative language may have very few exceptions to construed aggregation based coding of number neutral uncountable (collective) terms while another language has many exceptions. Tables 4.5 and 4.6 exemplify terms in Hejazi/Modern Standard Arabic and Welsh that are arguably well defined for construed aggregation, and yet are not collective/singulative marked terms.[15] The picture painted is that while Arabic has very few exceptions in the

[14]Of course, there is a possibility that specific diachronic reanalysis of terms has taken place. For example, PEAS being a plural-marked reanalysis of mass PEASE.

[15]Note, Table 4.6 does not include Welsh faux singulatives, which are also defined for construed aggregation and yet are singular/plural terms (see §2.1.5).

categories of fruits and insects, this exception class is decidedly larger in Welsh. With this, any construed aggregation approach to the count/mass distinction must take construed aggregation to be a necessary but not a sufficient factor associated with the lexicalist pre-syntactic generative component.

Mass Terms	**Count Terms**
amaranth, barley	peas, beans
buckwheat, bulgar	oats*, lentils
corn, farro	seeds
freakeh, millet	
orzo, popcorn	
quinoa, rice	
rye, spelt	
sugar, urad	
wheat, yeast	

* = Plurale tantum/singular form rarely used

Table 4.4: English granular aggregate terms

	Arabic	**Gloss**
Singular/ Plural	niktaariin → niktaariin-aat	nectarine → nectarines
	butghuuth → baraghiith ▼	flea → fleas
	3nkabuut → 3nakib ▼	spider → spiders
	3qrab → 3qarib	scorpion → scorpions
	dabbar → dabaabiir ▼	wasp → wasps
	ṣirṣar → ṣaraṣiir	cockroach → cockroaches
	khunfusa’ → khanaafis	beetle → beetles
	jundub → janadib	grasshoppers → grasshopper

▼ The equivalent term is collective/singulative in Welsh

Table 4.5: Exceptions to the Arabic collective class

	Count term	Gloss
	almon → almon-au ▼	almond → almonds
	pecan → pecan-au ▼	pecan → pecans
	olif → olif-au ▼	olive → olives
	nectarîn → nectarin-au	nectarine → nectarines
	afocado → afocado-s ▼	avocado → avocados
	mango → mango-au	mango → mangoes
	ciwi → ciwïod ▼	kiwi → a kiwi
	afal → afal-au ★ ▼	apple → apples
	oren → orenn-au ▼	orange → oranges
	lemon → lemon-au ▼	lemon → lemons
	tomato → tomato-s ▼	tomato → tomatoes
	pupur → pupr-au ▼	peppers → a pepper
	banana → banana-s ▼	banana → bananas
Singular/ Plural	pomgranad → pomgranad-au ▼	pomegranates → a pomegranate
	grawnffrwyth → grawnffrwyth-au ★	grapefruit → grapefruits
	pîn-afal → pîn-afal-au ▼ ★	pineapple → pineapples
	ciwcymbr → ciwcymbr-au ▼ ▼	cucumber → cucumbers
	corbwmpen → corbwmpenn-i ★ ▼	courgette → courgettes
	melon → melon-au ▼	melons → a melon
	dyfrfelon → dyfrfelon-au ★▼	watermelons → a watermelon
	maro → maro-s	marrow → marrows
	pompiwn → pompiyn-au ▼	pumpkin → pumpkins
	miltroed → miltroed-iaid ★	millipede → millipedes
	sgorpion → sgorpion-au	scorpion → scorpions
	mosquito → mosquito-s ▼	mosquito → mosquitos
	locust → locust-iaid ▼	locust → locusts
	glöyn byw → glöynn-od byw ▼ ★	butterlfy → butterflies
	pili-pala → pili-palod ★ ▼ (S)	butterfly → butterflies
	rhesin-en → rhesin-s	raisin → raisins
	wnion-yn → nion-od ▼	onion → onions
	tat-en → tat-ws	potato → potatoes
	trog-en → trog-od ★	tick → ticks
Faux Singulatives	malw-en → malw-od (N) ★▼	snail → snails
	gwlith-en → gwlith-od ★▼	slug → a slug
	sionc-yn gwair → sionc-od gwair ★	grasshopper → grasshoppers
	chwil-en ddu → chwil-od duon ★	cockroach → cockroaches
	chwil-en → chwil-od ★	beetle → beetles

★ The term is *not* borrowed.

▼ The equivalent term is collective/singulative in Arabic

Table 4.6: Exceptions to the Welsh collective class

4.2.3.3 Grimm (2012, 2018)

As Link (1983) is the seminal work for mereological approaches to nominal number, so too is Grimm (2012a,b, 2018) the seminal work for the construed aggregation approach to nominal number in singulative languages. For Grimm (2012b), the pre-syntactic generative component is modelled through the `THE SCALE OF INDIVIDUATION` in Figure 4.3. The Scale of Individuation comprises of (at least) four `INDIVIDUATION TYPES`. The `LIQUIDS AND SUBSTANCES` individuation type captures entities which have no individuation, such as water. The `GRANULAR AGGREGATE` individuation type captures entities that typically touch, such as grains of sugar. The `COLLECTIVE AGGREGATES` individuation type captures entities that are not touching, but are connected via function or aggregation, *i.e.* may come in groups, such as ants. Finally, the `INDIVIDUAL` individuation type captures entities that are not connected in any way via aggregation or function, such as dogs.

Liquids/substances $<$ Granular Aggregates $<$ Collective Aggregates $<$ Individuals

Figure 4.3: Grimm's Scale of Individuation

To formalise the The Scale of Individuation, Grimm uses a mereotopological approach to number.[16] The conditions for the individual individuation type are in (17), where the mereotopological concept of being maximally strongly self connected (MSSC) relative to a property P is used to capture that individuals have parts that are integrated wholes (with respect to a predicate).

[16] Grimm (2012b)'s account is not the only one to use mereotopological notions to define number categories with respect to collective/singulative languages. In more recent work, mereotopological approaches are used by Wągiel (2021a,b); Wągiel and Shlikhutka (2023a) in the analysis of Ukrainian singulatives. I choose to show Grimm (2012b,a)'s account for two reasons. First, Grimm should be properly acknowledged for being the pioneer of mereotopological methods in linguistics. Second, Grimm alone discusses the mapping from construed ontological concept to linguistic ontological nominal type. This being said, Wągiel (2021b) discusses some formal issues with Grimm's logical representation, specifically with how the transitive connection and cluster connection are defined. Despite this, I present Grimm's formulation almost verbatim, changing only the choice of terminology/symbols to be consistent within this thesis.

(17) `INDIVIDUAL INDIVIDUATION TYPE`

$\textsc{individual}(P) \rightarrow \forall x[P(x) \rightarrow \exists y[y \sqsubseteq x \wedge \textsc{mssc}(y,P)]]$

A predicate is INDIVIDUAL *if any individual that satisfies the predicate has a (not necessarily proper) part which is a* MSSC *individual.*

From Grimm (2012a, p.594)

To capture the conditions of the liquids/substances individuation, Grimm uses the mereotopological concept of being firmly connected (FC), and relativises it to a property as in (18).[17] Firm connection in this way captures the intuition that multiplicity is involved, *i.e.* that an instance of water strongly implies another instance of water.

(18) `SUBSTANCE INDIVIDUATION TYPE`

$\textsc{substance}(P) \rightarrow \forall x[P(x) \rightarrow \exists x'[P(x) \wedge x' \neq x \wedge \textsc{fc}(x,x')]]$

If P is a substance predicate then all individuals that satisfy P are firmly connected to a distinct individual of the same substance

From Grimm (2012a, p. 596), adapted

The descriptions Grimm (2012a,b) gives to capture the granular aggregate and collective aggregate individuation types are more involved. Two types of aggregation are modelled. To model aggregation where entities typically touch, Grimm uses the mereotopological `EXTERNAL CONNECTION` (EC) in (19), the idea being that granular aggregates are typically in touch.[18]

[17] Grimm uses the terminology `STRONG CONNECTION` for the same property. I follow Wągiel (2021b) in using Varzi (2007)'s term `FIRM CONNECTION` as not to confuse with the `STRONGLY SELF CONNECTED` property. This is simply a matter of terminology and has no bearing on the formalisms presented.

[18] As will be discussed in the chapter 5, external connection is not appropriate to model a series of aggregates due to the fact that two externally connected individuals either share a boundary, or one of the individuals is boundaryless. Either way, external connection requires two individuals to be inseparable.

(19) `EXTERNAL CONNECTION`

$\textsc{ec}(x,y) \overset{\text{def}}{=} \mathcal{C}(x,y) \wedge \neg\mathcal{C}(\textsc{int}(x), \textsc{int}(y))$

Two entities x and y are externally connected if they are connected, but their interiors do not overlap.

From (Grimm, 2012a, p.595), adapted

To model aggregation where entities do not typically touch but are sufficiently near, Grimm posits a new 'connection': `PROXIMATELY CONNECTED` (PC) as in (20). The proximate connection is made up of two parts. The first part is a standard distance function which takes two individuals as arguments and outputs a Euclidean measure of their distance.[19] The second part is a value $\boldsymbol{n}$ for the maximum distance two individuals may be for them to be considered sufficiently close to each other. The value of $\boldsymbol{n}$ is determined relevant to the predicate. The point of the flexible value of $\boldsymbol{n}$ is such that in different contexts, the threshold for proximity changes. For example, the threshold for closeness of ducks is not comparable to the closeness of ants. Explicitly, if two ducks are five meters apart, they be considered proximately connected, but if two ants are five metres apart, they may not be considered proximately connected.[20]

(20) `PROXIMATELY CONNECTED`

$\textsc{pc}(x, y) \overset{\text{def}}{=} d(x,y) \leq \boldsymbol{n}$

Two entities x and y are proximately connected if the distance between then is less than or equal to a contextually specified amount $\boldsymbol{n}$*.*

From Grimm (2012a, p.595)

With these definitions set, Grimm then defines the `TRANSITIVELY CONNECTED` (TC) and the `CLUSTER` (CLSTR) relations. The idea that underlies transitive connectedness

[19] See §3.6 for more on distance functions.

[20] Grimm (2012b)'s original formulation of proximate connectedness does not call for a predicate in the formalisation. In more recent renditions, such as Wągiel (2021b)'s alteration, a predicate is required in the formalisation.

is that two entities may be 'connected' through a series of like-entities that have the same connection relation. A pile of sand is said to be defined for transitive connectedness if it is made up of a series of externally connected sand entities. Similarly, a group of ants is defined for transitive connectedness if they are made up of a series of proximately connected ant entities.

(21) `TRANSITIVELY CONNECTED`

$\textsc{tc}(x,y, P, C, Z) \stackrel{\text{def}}{=} \forall z \in Z[P(z) \wedge (x{=}z_1 \wedge y = z_n) \wedge \mathcal{C}(z_1,z_2) \wedge \mathcal{C}(z_2,z_3) \ldots \wedge \mathcal{C}(z_{n-1},z_n)\ ]$ where $Z = \{\ z_1,\ z_2,\ldots.z_n\}$

Two entities x and y are transitively connected relative to a property P, a connection relation C, and a set of entities Z, when all members of Z satisfy P and x and y are connected through the sequence of z_is in Z.

From Grimm (2012a, p.598), adapted

The cluster relation is then defined for entities that refer to the mereological sum of transitively connected individuals. Grimm (2012b, 2018)'s original definition of CLUSTER is given in (22).

(22) `CLUSTER`

$\textsc{clstr}(x,P,C) \stackrel{\text{def}}{=} \exists Z[x = \bigsqcup Z \wedge \forall z,z' \in Z\ \exists Y[\textsc{tc}(z,z',P,C,Y)]]$

x is a cluster relative to a property P and a connection relation C iff x is a sum of entities falling under the same property which are all transitively connected relative to some set Y under the same property and connection relation.

From Grimm (2012a, p.598), adapted

With these definitions in place, Grimm (2012a,b) gives the general condition of an aggregate predicate as in (23). Under this definition, two types of aggregates can be defined, depending on whether the cluster is defined relative to an external connection relation (granular aggregates) or a proximately connected relation (collective aggregates).

(23) $\text{AGG}(P,C) \rightarrow \forall x[P(x) \rightarrow x \in \textsc{cluster}_c \cup \textsc{sum} \cup \textsc{mssc}\ ]$
If P is an aggregate predicate relative to a connection relation then all individuals that satisfy P have cluster, MSSC or sum reference
From Grimm (2012a, p. 598), adapted

With the formalisms set, Grimm (2012b, 2018) offers an account for the coding not only of collective/singulative terms in singulative languages, but also for object mass terms (for natural aggregate kinds) in number marking languages such as English. For Grimm, The Scale of Individuation is a series of nominal equivalence classes, and languages may assign different categories of nominal number to each individuation type on the scale. English is said to assign mass/zero coding to liquids and substance individuation types, while assigning singular/plural the collective aggregate and individual individuation types. Welsh is said to assign singular/plural coding to the individual individuation type only, while collective/singulative coding is assigned to the granular and collective aggregate individuation types. Though Grimm does not specifically mention any (dialects of) Arabic, the arguments made for Welsh can be assumed to hold for Arabic. The attested systems of Welsh, Arabic and English are demonstrated in Table 4.7, which is adapted from Grimm (2018, p. 26).

<table>
<tr><th>Language</th><th>Liquids / Substances</th><th>Granular Aggregates</th><th>Collective Aggregate</th><th>Individuals</th></tr>
<tr><td>Welsh</td><td>0</td><td colspan="2">0/Singulative</td><td>0/plural</td></tr>
<tr><td>Arabic</td><td>0</td><td colspan="2">0/Singulative</td><td>0/Plural</td></tr>
<tr><td>English</td><td colspan="2">0</td><td colspan="2">0/plural</td></tr>
</table>

Table 4.7: Crosslinguistic number according to Grimm's Scale of Individuation

While Grimm (2012b, 2018) takes The Scale of Individuation to be absolute for mapping individuation type to grammatical class, it is not the case that in this system that all *notional* aggregates are mapped to an aggregate individuation type. Grimm is careful to explain that The Scale of Individuation is not directly based on entities in the world, but rather it is based on *construals* of entities in the world. Following Stolz (2001), Grimm (2018, p.39) argues that it is not the literal gregarious nature of individuals that lead to collective/singulative coding, but rather the *conceptualisation* that certain individuals come in groups that leads to collective/singulative coding.

To capture this in a formal way, Grimm (2018, p.42) posits that an `ENTITY DESCRIPTION` lexicalises construals of the world. This entity description includes information pertaining to any construals of aggregation. The entity description is mapped to the relevant individuation type with respect to The Scale of Individuation, which is then in turn mapped to the relevant grammatical class in a given language. For example, entities in the world that are ducks have the entity description DUCK in English, HWYAID in Welsh, and BAṬ in Arabic. As ducks may very well be defined for aggregation (they come in groups and are defined for proximate connectedness), these entity descriptions are mapped to the collective aggregate individuation type. At this point, English, Welsh and Hejazi/Modern Standard Arabic differ in how they cut The Scale of Individuation as in Table 4.7. Therefore, English DUCK is mapped to singular/plural nominal class, while Welsh HWYAID and Hejazi/Modern Standard Arabic BAṬ are mapped to the collective/singulative class.

In this manner, Grimm (2018) is explicit in his assumptions for the connections between fundamental ontologies (entity-in-world), construed ontologies (entity description/individuation type) and linguistic ontologies (individuation type/grammatical class). A general picture of this process is demonstrated in Figure 4.4, which is adapted from a similar figure in Grimm (2018, p.42).

entity	entity-in-the-world1	entity-in-the-world2
entity description	noun 1	noun 2
individuation type	ind type 1 < ind type 2 <	···· < ind type m-1 < ind type m
grammatical class	Grammatical Class 1	Grammatical Class 2

Figure 4.4: Grimm's mapping of the levels of ontology.

With this mapping in place, let us now turn to Grimm (2012b)'s model of the Welsh nominal system. The entity description for GRAWN (*'grains'*) presumably contains an aggregation construal where entities are defined for the cluster relation relative to external connection. As such, GRAWN is mapped to the granular aggregate individual type, and then in turn is mapped to the collective/singulative class. Similarly, the entity description CACWN (*'wasp'*) contains an aggregation construal where entities are defined for the cluster relation relative to proximate connection, leading to mapping to the collective aggregate individuation type, and therefore mapping to the collective/singulative class. Finally, CATH (*'cat'*) has no aggregation construal, and so is mapped to the individual individuation type, which in turn is mapped to the singular/plural class. The semantics Grimm offers for GRAWN, CACWN, and CATH are in (24), (25), and (26), respectively.

(24) $\textsc{grawn} = \lambda \mathrm{x_o}[\mathrm{R}(\mathrm{x_o}, \text{grain}) \wedge x_0 \in \textsc{cluster}_{\text{extc}}]$

(25) $\textsc{cacwn} = \lambda \mathrm{x_o}[\mathrm{R}(\mathrm{x_o}, \text{wasp}) \wedge \mathrm{x_o} \in \textsc{cluster}_{\text{proxc}}]$

(26) $\textsc{cath} = \lambda \mathrm{x_o}[\mathrm{R}(\mathrm{x_o}, \text{cat}) \wedge \mathrm{x_o} \in \textsc{mssc}]$

Following this, Grimm gives the semantics in (27) for the Welsh plural and (28) for the Welsh singulative. The semantics are such that the plural will only combine with terms which denote MSSC individuals, outputting an equivalent of Link (1983)'s ^{c}P, *i.e.* a set of sums. This correctly rules out that collective terms may pluralise as they denote clusters. Similarly, the singulative morpheme combines only with those terms which denote clusters, outputting a set of MSSC individuals.

(27) $\textsc{plural} = \lambda \mathrm{P} \lambda \mathrm{x.P_{\textsc{mssc}}}[{}^{*}\mathrm{P(x)} - \textsc{mssc}(\mathrm{x})]$
$= \lambda \mathrm{P.P_{\textsc{mssc}}}[{}^{*}\mathrm{P(x)} \wedge \mathrm{x} \in \mathrm{SUM}]$
From Grimm (2012a, p.600), adapted

(28) $\textsc{singtive} = \lambda \mathrm{Q}\ \lambda \mathrm{x.Q_{cluster}}[\mathrm{x} \sqsubseteq \mathrm{Q} \wedge \mathrm{x} \in \textsc{mssc}]$
From Grimm (2012a, p.601), adapted

With these definitions, the plural will combine with both singular terms, *e.g.* CATH (cat.SG = *'cat'*) as in (26) to reveal CATHOD (cat.SG-PL = *'cats'*) and also to singulative terms, *e.g.* CACYN-YN (wasp.COL-SING = *'wasp'*) to reveal CACYN-YN-NAU (wasp.SG-PL = *'wasps'*). This state of affairs is unfortunate, as the Welsh individuating singulative does not pluralise, as shown in (29).[21]

(29) *Gwelodd Owain **hwyad-enn-au** yn yr ardd — Welsh
see.PST.3 Owain duck-SING-PL in DEF garden
Owain saw girls in the garden

Not only does Grimm (2012b)'s formal account outright predict ungrammatical structures in Welsh, the pre-syntactic generative component modelled as in Figure 4.4 is theoretically challenging. There is no mechanism to stop non-aggregate construals from being mapped to aggregate individuation type. As the denotational space of an aggregate includes the union of clusters, sums, and MSSC individuals as in (23), there is no immediate reason why a MSSC entity description with no construal of aggregation should not be mapped to it. This is perhaps an easy fix if we allow only entity descriptions with an aggregation construal to map to aggregation individuation types. This would systematically block, for example, the entity description BEAR from being mapped to an aggregate type.

There is one further issue which applies not to just Grimm's system, but is challenging for any lexico-meaning account: how much crosslinguistic variation is expected? And how can we account for exceptions to the rule? Consider that olives and wasps are potentially locally unmarked concepts, and so prime candidates for collective/singulative coding. Yet, Welsh OLIF (olive.SG = *'olive'*) and Arabic DABBAR (wasp.SG = *'wasp'*) are count terms. Under a Grimm style analysis, we would have to conclude that the entity descriptions OLIF and DABBAR have no aggregation construal, and so are mapped to the individual individuation type, and then subsequently mapped to the singular/plural class. Why should this be so?

[21] In defence of Grimm (2012b)'s work, there is an assumption that the Welsh singulative does pluralise, using the GRONYNNAU paradigm as evidence. See §2.3.2 for further discussion of the absence of the Welsh plural of the singulative.

We could perhaps argue a la Wierzbicka (1988) that olives have a higher degree of individuation as they have stones in them and so are typically eaten one-by-one, perhaps over a long period of time. Similarly, wasps have high individuation as they are stinging insects, and so individual interaction is seen at the level of a stinging event. The problem is that this style of argumentation would have to hold for *every* term that has escaped collective/singulative coding despite the individuals being defined for aggregation. It is not clear that such an account could be so easily given. Cherries, like olives, have stones in them and are often eaten one by one but CEIRIOS (cherry.COL = *'cherry/ies'*) is a collective/singulative term in Welsh. Why should interaction factors behave differently so that the entity description CEIRIOS is mapped to the collective aggregate individuation type and so becomes collective/singulative, but not OLIF? A similar story can be said for Hejazi/Modern Standard Arabic NAHAL (bee.COL = *'bee(s)'*). Why should NAHAL be mapped to the collective aggregate individuation type, but not DABBAR? This style of argumentation is also crosslinguistically uncomfortable, as a property may be collective/singulative in one language, but not another, as shown in Table 4.8.

Welsh	**Arabic**	**Gloss**
singular/plural	**collective/singulative**	
olif → olif-au	zaytuun → zaytuun-ah	olive(s)
lemon → lemonn-au	leymuun → leymuun-ah	lemon(s)
afal → afal-au	tufaaḥ → tufaaḥ-ah	apple(s)
buwch → buch-od	baqar → baqar-ah	cow(s)
cyw→ cyw-ion	dajaaj → dajaaj-ah	chicken(s)
collective/singulative	**singular/plural**	
pryfed cop → pryf cop-yn	3nkabuut → 3nakib	spider(s)
chwain → chwann-en	butghuuth → baraghiith	flea(s)
cacwn → cacyn-en	dabbar → dabaabiir	wasp(s)
moch → moch-yn	khinziir → khanaaziir	pig(s)

Table 4.8: Variation of the collective/singulative terms across Welsh and Arabic

How should the pairs in Table 4.8 be explained? Is it the case that we should argue that Arabic speakers and Welsh speakers interact with certain entities in different ways such that their construals are affected, ultimately leading to crosslinguistic variation? Indeed, for Grimm such crosslinguistic variation is unproblematic as:

> there is much room for disagreement among languages and cultures as to whether an entity should be described in one fashion or another.
> (Grimm, 2018, p.26).

While we may be able to explain away each and every term which has escaped collective/singulative coding, this style of reasoning runs into the same criticisms that Wierzbicka received from Palmer (1990), *i.e.* these are speculative assumptions where, without rigid experimentation, it is impossible to know how cultures and people think about certain entities. Going down this route runs into danger of being forced, ad-hoc, and (without knowing what cultures think) unfalsifiable.

It may seem that some of these criticisms are overstating the point. From Middleton et al. (2004)'s experimental evidence, it is shown that interaction with individuals rather than the group can lead to mass coding in experimental contexts. But it seems ad hoc and unreasonable to assume that all exceptions to collective/singulative coding are due to interaction effects. We are then at a crossroads with how interaction of construed ontological construals and linguistic coding of nominal number categories. After all, Grimm (2012b)'s account captures one crucial fact that Chierchia (2010b) and Landman (2020)'s accounts do not: in number marking and singulative languages, while there are number neutral uncountable terms which are not vague nor overlapping, there are no number neutral uncountable terms (that I am aware of) that cannot be argued to be defined for aggregation, *i.e.* denoting an aggregate, swarm, collection *etc.*. In this way, Grimm captures that construed aggregation seems to be necessary, but not sufficient for collective/singulative coding in number marking languages.

There is one final point with regard to Grimm's work that is vaguely unsatisfying. The account only affords analysis for singulative languages and number marking languages. It is not clear how it may extend to generalised classifier languages, such as Mandarin Chinese. The singulative marker in Welsh and Arabic has the same underlying primitive properties as the general classifier: it derives a singular countable term from a number neutral uncountable term (see §2.2.3). Grimm's account does not capture this basic similarity. The singulative semantics he offers presupposes that it applies only to nominal terms which denote clusters. It is obviously not the case that all nouns in *e.g.* Mandarin Chinese denote clusters. Therefore, under a Grimm (2012b)-style approach, singulatives and general classifiers hail from different sources.

4.2.4 Lexicalism summary

Lexico-meaning approaches to nominal number are alluring as they offer parsimony between construed ontologies and linguistic ontologies. Humans have a pre-conceived notion of the substance/object distinction, and lexicalist approaches shine in their ability to capture something seemingly obvious; this pre-conceived substance/object distinction is relevant for natural language. The mapping property captures that there is no standard for ONE WATER or ONE SNOW, and so languages will never lexicalise these concepts as countable terms. The problem that all lexico-meaning approaches face is the proper treatment/mapping of number neutral uncountable terms. As there is no clear set of number neutral uncountable terms across languages, lexicalist accounts must either rely on a.) an element of arbitrariness of the generative system as in Chierchia (2010b) and Landman (2020), or b.) increasingly use unfalsifiable claims of culture variation as in Grimm (2012b, 2018).

Even if the lexico-meaning pre-syntactic generative component for number neutral uncountable terms in number marking and singulative languages is identified, there must then be some mechanism that allows generalised classifier languages to completely sidestep any proposed construed ontological/linguistic ontological mapping. In Chierchia (2010b)'s account, canonical mass terms are subject to the pre-syntactic generative component, but fake mass terms in English, and all Mandarin Chinese terms (modelled as kinds), escape any construed ontological based source of 'mass' coding. In Landman (2020)'s account, the mapping from concept to number neutral uncountable for English is based in vertical overlap, while for Mandarin Chinese it is completely arbitrary. In Grimm (2012b)'s account, there are four individuation nominal equivalence classes, with no mention of how a generalised classifier language like Mandarin Chinese may fit into the system.

Though lexicalist accounts must rely on an increasing number of mechanisms to derive the correct count/mass coding of nominal terms, these mechanisms are not inherently problematic. Indeed, there is some level of arbitrariness required for nominal coding to explain *why* all terms are number neutral uncountable in generalised classifier languages, but only some are in number marking languages and singulative languages. Similarly, if a lexicalist approach is correct and nominal categories are inherently linked to some construed ontological feature, then variation between cultures and people must affect nominal coding, for if we expect coding for like-terms to be identical across all language we must posit the thoughts/perceptions of all cultures are identical. Therefore, the issue with lexico-meaning approaches is not that arbitrariness or reliance on cultural factors interacts with the pre-syntactic generative component. Rather, the issue is the *extent* to which arbitrariness and cultural factors can affect nominal coding: at what point is it unreasonable?[22]

[22]There is a parallel discussion here regarding arbitrariness and linguistic representations in other domains, such as grammatical gender. In many languages grammatical gender assignment can be derived from some social/conceptual notion of gender (see *e.g.* the distribution of the feminine sound plural in Arabic in Table 2.3), but in other cases grammatical gender has no clear link to social/conceptual gender (*e.g.* all collectives are masculine in Arabic, but not in Welsh). The difference between discussions of arbitrariness in gender assignment and arbitrariness in collective/singulative assignment is, I surmise, that individuals in the world will always have construals of individuation, *i.e.* associations of whether it naturally occurs in groups, however those same individuals will not necessarily have associations with social/conceptual gender construals. I suggest that when there is no association of gender construals with a concept/individual, then other means will assign grammatical gender (*e.g.* phonological shape, arbitrariness). However, in the case of number (*i.e.* collective/singulative) assignment, there are necessarily relevant construals that influence the system. Therefore, we may expect there to be comparatively less arbitrariness of linguistic representation in the number assignment domain than in the gender assignment domain.

4.3 Constructionism

An alternative approach to the theory of nominal number, famously spearheaded by Borer (2005a) (see also Sharvey, 1978), would be to abandon any link between default nominal coding and construed ontologies. In essence, a constructionist approach to nominal number implicitly denies a Quineian split of 'divided' and 'non divided' lexical items. Under the constructionist approach, no term comes with 'individuation built in', and therefore there is no pre-syntactic generative component that applies at the lexical level. Instead, a constructionist account posits that the lexicon is impoverished with respect to number, and so all number distinctions are generated syntactically. The consequence of this is that any term may be used in a count or mass way. In this manner, crosslinguistic variation is simple: all languages have the same basic tools, and there is no need to posit that construed ontological judgements predict nominal number. While there have been many advancements to the constructionist approach to nominal number over the years (see Borer and Ouwayda (2021)), I present the motivation in light of Borer (2005a)'s original wisdom.

Note, while I am discussing constructionist accounts where there are no count and mass terms (only count and mass uses), I will continue to use familiar terminology of `COUNT`, `MASS` and `NUMBER NEUTRAL UNCOUNTABLE` to disambiguate intended readings.

4.3.1 Rejecting the mapping property

If one is to follow a constructionist approach to nominal number, this comes with an implicit denial of the mapping property: if lexical items are neither inherently mass nor count, then there is no reason to map construed ontologies to linguistic ontologies at the nominal level with respect to 'default' number. Assuming this, there is nothing that blocks a certain syntactic use of mass or count structure. This does not mean that the construed ontological level is irrelevant to linguistic representations, but rather that the role of perceptual reality is to modulate the use/interpretation of structures in context. This being the case, perceptual realities are not necessarily the deciding factor for interpretation in context. As Hagit Borer puts it:

> The distinction, I will argue, is strictly a grammatical one, and although our perception of the world might lead us to prefer some structural combination over another in conjunction with particular concepts, consideration of world knowledge, as has already been suggested, can be routinely overridden by grammatical factors.
> (Borer, 2005a, p.17)

For Borer (2005a), the argument that underlies this approach is one of nominal flexibility. Borer argues that every term can be used in a mass or count way *e.g.* grinding, portioning *etc.*, as seen in (30) and (31). Therefore, according to Borer, there is no reason to take mass and count features as lexically given.

(30) Count use of 'mass' and 'count' terms

a. Edward bought two cats — English

b. Edward bought two waters

(31) Mass use of 'mass' and 'count' terms

a. Edward put water in the soup — English

b. Edward put cat in the soup

4.3.2 Acategorical roots

To formalise these intuitions, Borer (2005a) states that nPs begin life as an acategorical root that is neither mass nor count, but denotes only 'stuff'. Countability features are derived via (the lack of) the projection of a divisional head, ***`div`***. In the case that ***`div`*** is not projected, the stuff reading of the root will abound. In the case that ***`div`*** is projected, the root meaning will be 'portioned out' and the count reading will arise. The portions derived via ***`div`*** may constitute canonical pre-theoretical individuals, or they may not. The point is, every root will need to be portioned via ***`div`*** so that the individuals are revealed for counting purposes. The counting feature in Borer's system is located in a head dominating *div*, namely #, which hosts cardinals in its specifier. In the case that both # and *div* are present in a structure, then # counts the portions created by *div*. The basic architecture of Borer (2005a)'s system is demonstrated in Figure 4.5.

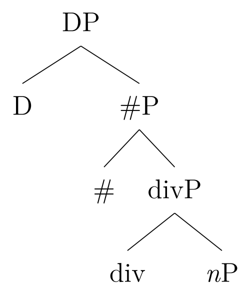


Figure 4.5: Borer's exoskeletal structure for nominal phrases

Borer states that the content (or assigned range) of *div* varies intra- and cross-linguistically. For generalised classifier languages, *div* may be realised as an independent morph, such as a classifier, as in Figure 4.6 for Mandarin Chinese. It is by this measure Borer derives ungrammaticality of numerals with bare terms in generalised classifier languages: in the absence of *div*, there are no portions to count. In English, *div* is realised as a phonologically abstract head, in which case head movement of the nP head is triggered, and *div* is realised as a suffix. This is said to be the case of 'plural' marking, as in Figure 4.7. In this vein, the English 'plural' is said to have the same function as a classifier. The absence of 'plural marking' then implies *div* is not present in the structure, and so portions cannot be measured out, deriving the ungrammaticality of *e.g.* TWO CAT. Finally, for language which can combine numerals with singular nominals (as in Turkish or Hungarian), Borer posits that *div* and # may form a complex, where the same morpheme (a numeral) has cyclically raised through each head and thus behaves as both a divider and a counter, as in Figure 4.8.

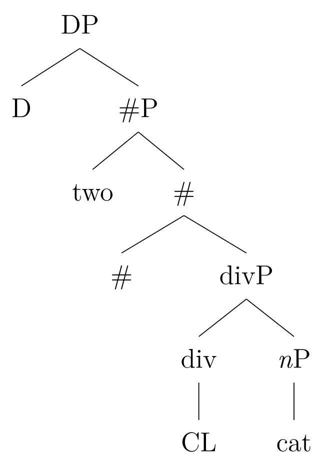


Figure 4.6: Borer's nominal structure for Mandarin Chinese

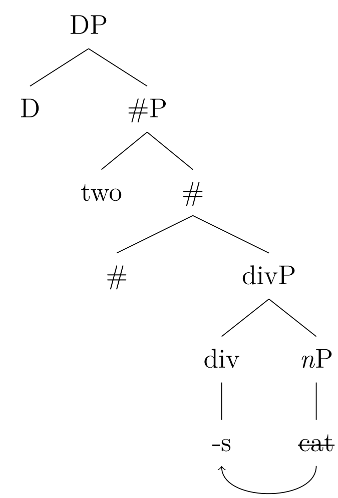


Figure 4.7: Borer's nominal structure for English

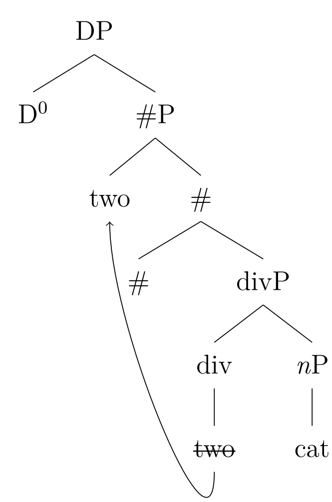


Figure 4.8: Borer's nominal structure for Turkish

Taking Borer (2005a)'s proposal seriously, two interesting consequences arise. First, with respect to nominal flexibility, the interpretation of coerced stuff mass (*e.g.* TWO WATERS) arises because the root WATER$^{\sqrt{c}}$ has been portioned out via -S in ***div***, which allows for its subsequent combination with cardinals. Second, as Borer (2005a) explains, the reason plural marking and classifiers do not co-occur is because they are in complementary distribution and their role is one-and-the-same, *i.e.* division in ***div***.

The benefits of a constructionist approach to number are promising. Though the specific content of ***div*** varies across languages, what emerges is an exoskeletal structure that is powerful enough to derive the syntactic and semantic behaviours of different languages in a structural way without the need to appeal to spurious conversation concerning cultural or construed ontological explanations. As such, there is no construed ontological/linguistic ontological mapping arbitrariness within the system, either at the crosslinguistic or language specific level. Further, as count and mass terms may always be derived, the absence of a given structure is not due to illegal derivation, but rather usefulness in context.

While Borer's constructionist approach is powerful and promising, it is not without its theoretical flaws. As has been pointed out by Rothstein (2010, 2017); de Vries and Tsoulas (2021); Bale and Barner (2009) *a.o.*, Borer's account, as it stands, is unable to distinguish between number neutral uncountable reference and stuff mass reference. This is because mass terms are characterised by a lack of ***div***, which in turn equates meaning with 'stuff'. There is no clear reason as to why *e.g.* English WATER and Mandarin Chinese SHUI (*'water'*) should refer to *stuff* while English FURNITURE and Mandarin Chinese MAO (cat.NNU = *'cat(s)'*) refer to (sums of) individuals if they are all characterised by the lack of ***div***.

4.3.2.1 Mathieu (2012, 2013, 2014)

Accounting for collective/singulative systems in a constructionist framework is not a simple task. The first attempt I am aware of was by Mathieu (2012a, 2013, 2014) who noted that Borer (2005a)'s account is not fully compatible with the Arabic collective/singulative system. Mathieu correctly observes that collective terms are number neutral uncountable, as shown in the original example in (32). Therefore under a Borer-style approach should be associated with the lack of a ***div*** projection. As the singulative marker is a marker of individuation, as shown in the original example in (33), Mathieu associates singulative marking with ***div***.[23] While this analysis correctly captures that singulative marking is required for counting, Mathieu notes that the Arabic singulative may pluralise, as in the original examples in (34) and (35). As Borer's account takes pluralisation and classification to be in complementary distribution, the plural of the singulative is not predicted.[24]

(32) Malik shaaf **baṭ** fi al-hadiqa — H. Arabic
Malik see.PST.M duck in DEF-garden
'*Malik saw a duck / ducks in the garden*'

(33) Malik shaaf **baṭ-at** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 duck-SING in DEF-garden
'*Malik saw a duck in the garden*'

(34) Malik shaaf **khams-ah baṭ-a-at** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 five-F duck-SING-PL in DEF-garden
'*Malik saw five ducks in the garden*'

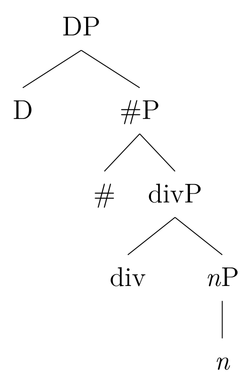
(35) Malik shaaf **baṭ-a-at** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 duck-SING-PAU in DEF-garden
'*Malik saw a small number of ducks/trees in the garden*'

[23] See also the works of Abdelkader Fassi Fehri (2016; 2020) who also places the singulative under a form of ***div***.

[24] Note, Borer (2005a)'s account does not predict that generalised classifier languages should not have plural markers at all, but rather that they should not co-occur with classifiers. There are arguments that Mandarin Chinese, Japanese, and Korean have plural markers (-MEN, -TATI, -TUL, respectively), but they do not typically occur with classifiers. Having noted this, these so-called plural markers are not actually plural markers in the sense that English -S is a plural marker, as they have meanings relating to associativity and distributivity. See Cheng and Sybesma (1999); Kim (2008); Li (1999); Kang (1994); Iljic (1994) for further details.

There are two routes that can be taken to explain the Arabic plural of the singulative. The first is that the plural of the singulative is not a functional plural at all, but simply an agreement marker. This is the route taken by Borer and Ouwayda (2010, 2021); Ouwayda (2014).[25] Yet, this is problematic as the Arabic plural of the singulative can appear without numerals as in (35), making an agreement-based approach unlikely.

The second explanation for the Arabic plural of the singulative, taken by Mathieu (2012a, 2013, 2014), is that the plural of the singulative is a # head (presumably triggering raising), and as such has counting but no individuating features.[26] The # head may then appear bare, or as cardinals appear in the specifier of #, there is no problem in the emergence of numerals with the plural of the singulative. The schematic in Figure 4.9 emerges as the possible 'flavours' of division under Mathieu's approach. Figure 4.9 is adapted from Mathieu (2014, p.16-17).



- Contents of #
  - Plural of the singulative
- Contents of `div`
  - Plural (English, Arabic sound plural)
  - Numeral Classifier (M. Chinese)
  - Numerals (Turkish)
  - Singulative (Arabic)
    - Gender Shift
    - Diminutive

Figure 4.9: Mathieu's system of nominal number

[25] Strictly, the plural of the singulative in *e.g.* Borer and Ouwayda (2021) is a split-DP structure similar to Mathieu (2012a)'s suggestion but the key difference is that in Borer and Ouwayda (2021) the split structure must be licensed by a cardinal.

[26] See also Fassi Fehri (2016, 2020) for an account in a similar spirit.

As Mathieu's proposal is built upon the primitives set out by Borer (2005a), the proposal enjoys the same benefits as Borer's account. Namely, Mathieu's Borer-inspired account fares well with dealing with crosslinguistic variation in that it allows for a crosslinguistic unification of mass/count structures *i.e.* some (un)divided nP is the locus of the distinction. Mathieu (2012a)'s account also boasts theory specific benefits. Mathieu's version of constructionism explicitly connects the similar functions of classifiers and singulatives in such a way that their syntax and semantics are akin (as ***div*** heads), a welcome feature of the system not readily apparent in for example Grimm (2012b)'s lexicalist approach.

As Mathieu's account enjoys the same benefits as Borer (2005a)'s account, it too faces similar issues. The (lack of) stuff mass interpretations for certain terms (*e.g.* FURNITURE/MAO) is still not easily explained, and the issue is compounded by the collective/singulative class of nominals. The example in (36a) shows that for Welsh, the lack of morphological number marking for 'count term' LLEW (lion.SG = *'lion'*) results in a stuff mass interpretation of lion-meat. Yet, in the same context the lack of morphosyntactic number marking for 'collective' terms MOCH (pig.COL = *'pig(s)'*) in (36b) does not result in a stuff mass reading. Rather, the interpretation of MOCH in this context is of rather unusual scene where whole pigs are in the soup, perhaps quite small to even fit in the bowl. This difference in meaning is not predicted as both MOCH and LLEW are structurally identical under this approach: a lack of ***div*** should result in nPs denoting 'stuff'. A similar situation is seen in Hejazi Arabic, though in this case the lack of morphosyntactic number marking does not lead to a stuff mass interpretation for either 'count' or 'collective' terms, as in (37a) and (37b), respectively. Again, this is not predicted if the structure lacks ***div*** and the nP simply denotes unportioned stuff.[27]

(36) a. Mae yna **lew** yn y cawl — Welsh
be.3 there lion in DEF soup
'*There is lion meat in the soup*'

[27] I return to this issue in more detail in chapter 7.

b. Mae yna **moch** yn y cawl — Welsh
be.3 there pig.COL in DEF soup
'*There are pigs in the soup*'
NOT: '~~*There is pig meat/flesh/stuff in the soup*~~'

(37) a. fi **asad** fi shurbat-ii — H. Arabic
in lion in soup-my
'*There is a lion in my soup*'
NOT: '~~*There is lion meat/flesh/stuff in the soup*~~'

b. fi **baqar** fi shurbat-ii — H. Arabic
in cow.COL in soup-my
'*There are cows in my soup*'
NOT: '~~*There is cow beef/meat/flesh/stuff in the soup*~~'

A second conceptual flaw that is specific to Mathieu (2013, 2014)'s account is that it does not rule out nPs combining with the 'wrong' `div` head. There is nothing in the theory that stops the singulative morpheme applying to 'non-singulative' roots, and then (in Arabic) subsequently combining with the a higher plural, as in (38) and (39). Of course, one may argue for some vocabulary insertion rule such that `div` is covert for 'count terms' and overt for 'singulative terms'. A version of this proposal is found in Dali and Mathieu (2021b); Mathieu and Dali (2021); Dali (2020), which I return to shortly.

(38) ASAD$^{\sqrt{c}}$ → *'asad-ah → *'asad-a-at — H. Arabic
lion.ROOT → lion-SING → lion-SING-PL

(39) LLEW$^{\sqrt{c}}$ → *llew-en — Welsh
lion.ROOT → lion-SING

Not only does Mathieu (2012a, 2013, 2014)'s account struggle with conceptual fact, there is unfortunately one glaring issue when considering the Welsh collective/singulative system specifically. Welsh is similar to Turkish in that while it has plural markers as in (40), the combination of numerals and nominal terms results in singular agreement as in (41). Under both Borer (2005a) and Mathieu (2012a, 2013, 2014)'s accounts, the Welsh numeral should merge in *div* as a divider, and raise to # for counting. However, if the singulative is generated under *div*, then Welsh singulatives and numerals should be in complementary distribution, contrary to fact, as shown in (42).

(40) Gwelodd Owain **ferch-ed** yn yr ardd — Welsh
see.PST.3 Owain girl-PL in DEF garden
'*Owain saw girls in the garden*'

(41) Gwelodd Owain **bum merch** yn yr ardd. — Welsh
see.PST.3 Owain five girl in DEF garden
'*Owain saw five girls in the garden*'

(42) Gwelodd Owain **bum hwyad-en** yn yr ardd — Welsh
see.PST.3 Owain five duck-SING in DEF garden
'*Owain saw five ducks in the garden*'

4.3.2.2 Dali and Mathieu (2021)

In a series of related work, Dali and Mathieu offer a slightly different constructionist approach to collective/singulative systems (Dali, 2020; Dali and Mathieu, 2020, 2021a,b; Mathieu and Dali, 2021).[28] In the rendition in Dali and Mathieu (2021b), lexical roots are taken to be underspecified for number and category and must combine with a classificatory head ***n*** which structures the root as a join-semi-lattice. The ***n*** head is underdetermined for count/mass status. Like in Borer's account, 'division' (or lack thereof) generates countability distinctions. In this system, the

[28]The earlier accounts of Fassi Fehri (2004, 2012) also attempt to capture the collective/singulative distinction using a features based approach, though they differ in their constructionist/lexicalist status. The Fassi Fehri (2012) account is similar to Dali and Mathieu (2021b) in that it is a constructionist account which places features in the divisional head, though the account also has the features manifested on roots, essentially making them lexically given.

num head takes the role of *div* where in the case it is projected, the nominal is count, when it is absent, it is mass (like in Mathieu's earlier account, collectives are subsumed under mass interpretations, with no projection of *num*).

Dali and Mathieu (2021b) take the semantics of *num* to be associated with Martí (2020)'s interpretation of Harbour (2014)'s number features. In essence, the features on *num* filter the semilattice structure imposed by *n* to reveal different mereological structures of elements and their sums. The $[\![\pm atomic]\!]$ feature in (43) determines whether the projection of *num* refers to atoms $[\![+atomic]\!]$ or pluralities $[\![-atomic]\!]$. The $[\![\pm minimal]\!]$ feature in (44) determines whether the *num* projection has elements with parts $[\![-minimal]\!]$ or elements without parts $[\![+minimal]\!]$. The $[\![\pm additive]\!]$ feature in (45) determines whether the *num* projection has for any two of its members their join $[\![+additive]\!]$ or not $[\![-additive]\!]$.[29]

(43) $[\![\pm atomic]\!]$

$$[\![+atomic]\!] = \lambda\mathrm{P}.\ \lambda\mathrm{x}.\ \mathrm{P(x)} \wedge \mathrm{atom(x)}$$
$$[\![-atomic]\!] = \lambda\mathrm{P}.\ \lambda\mathrm{x}.\ \mathrm{P(x)} \wedge \neg\mathrm{atom(x)}$$

(44) $[\![\pm minimal]\!]$

$$[\![+minimal]\!] = \lambda\mathrm{P}.\ \lambda\mathrm{x}.\ \mathrm{P(x)} \wedge \neg\exists\mathrm{y}\ \mathrm{P(y)} \wedge \mathrm{y} \sqsubset \mathrm{x}$$
$$[\![-minimal]\!] = \lambda\mathrm{P}.\ \lambda\mathrm{x}.\ \mathrm{P(x)} \wedge \exists\mathrm{y}\ \mathrm{P(y)} \wedge \mathrm{y} \sqsubset \mathrm{x}$$

(45) $[\![\pm additive]\!]$

$$[\![+additive]\!] = \lambda\mathrm{P}.\ \lambda\mathrm{x}.\ \mathrm{Q(x)} \wedge \mathrm{Q}\sqsubset \mathrm{P} \wedge \forall\mathrm{y}\ \mathrm{Q(y)} \rightarrow \mathrm{Q}(x \sqcup y)$$
$$[\![-additive]\!] = \lambda\mathrm{P}.\ \lambda\mathrm{x}.\ \mathrm{Q(x)} \wedge \mathrm{Q}\sqsubset \mathrm{P} \wedge \neg\forall\mathrm{y}\ \mathrm{Q(y)} \rightarrow \mathrm{Q}(x \sqcup y)$$

In this system, *num* may have multiple features associated with it: the choice is parametric and as such the recipe of feature bundles differs crosslinguistically. Therefore, the 'filtering' operations applied to the semilattice imposed by *n* is subject

[29] Q is taken to be a free variable in this system, though Martí (2020) and Dali and Mathieu (2021b) point out that the variable could be existentially bound.

to crosslinguistic variation. Dali and Mathieu provide the semantics for singular count nominals (in Arabic) as in (46), where the $[\![+atomic]\!]$ feature ensures that the derivation after the application of ***num*** includes refers to atoms, while the $[\![-additive]\!]$ feature ensures there are no joins in the derivation. This specific recipe of features in ***num*** 'filters' the semi-lattice structure imposed by ***n*** such that the output only has reference to the atomic elements (of the root concept).

(46) Semantics of singular terms (in Arabic).

$$[\![-additive]\!]([\![+atomic]\!]([\![nP]\!]))$$
$$= \lambda \mathrm{x.Q(x)} \wedge \mathrm{Q} \sqsubset([\![+atomic]\!]([\![nP]\!])) \wedge \neg\sqsubset \mathrm{y\ Q(y)} \rightarrow \mathrm{Q(x} \sqcup \mathrm{y)}$$

From Dali and Mathieu (2021b, p.40)

In Dali and Mathieu (2021b)'s system, the only difference between singular and singulative terms is that 'collective' terms have a $[\![+collective]\!]$ class feature on the classificatory ***n*** head. The $[\![+collective]\!]$ feature offers no semantic consequences, only a context for which an allosemic rule derives singulative marking as in (47). Therefore, singulative marked terms are not syntactically or semantically different from singular terms, but are simply a case of inverse marking. The minimal differences for singular and singulative terms are shown in the derivations in Figure 4.10 and Figure 4.11.

(47) Allosemic rule for singulative terms in (Tunisian) Arabic

$$[+\mathrm{fem}] \leftrightarrow \boldsymbol{num}_{\ [+\mathrm{atom};\ -\mathrm{additive}]} \ \text{———}\ \boldsymbol{n}_{[\![+collective]\!]}$$

From Dali and Mathieu (2021b, p.33), adapted

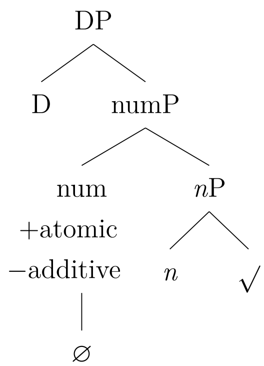


Figure 4.10: Dali and Mathieu's structure for singular count terms

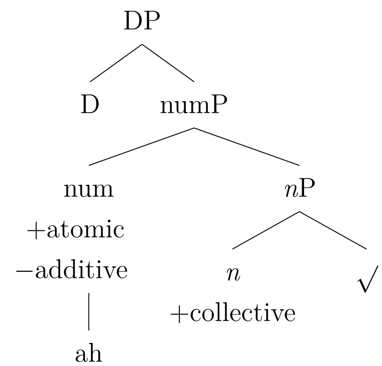


Figure 4.11: Dali and Mathieu's structure for singulative terms

Dali and Mathieu's proposal for plural marking is similar to Borer (2005a) in spirit as singular and plural marking is in complementary distribution, as all number features are in ***num***. In the simplest case for plural marking, a ***n*** with a complex bundle feature $[\![-atomic, +additive]\!]$ projects over $[\![nP]\!]$ as in Figure 4.12, ensuring there is no atomic reference. This is the semantics of the exclusive plural where the $[\![+additive]\!]$ feature ensures reference to sums, but the $[\![-atomic]\!]$ feature ensures no reference to singularities. For the plural of the singulative, the analysis is more involved. Dali and Mathieu adopt a split-DP hypothesis of nominal number where number features may be distributed across the nominal spine, similar in spirit to Mathieu's earlier work. In the case of the plural of the singulative, a secondary merger of ***n*** is proposed which 'resets' the semi-lattice structure so computation with a second ***num*** may start afresh. The semantics for the plural of the singulative emerge in this second projection of ***num***. Dali and Mathieu claim that the second projection of ***num*** in Arabic is necessarily associated with the complex feature bundle $[\![-atomic, -additive]\!]$, as the bare plural of the singulative is always paucal. In this case, the $[\![-atomic]\!]$ feature ensures exclusive plurality, and the $[\![-additive]\!]$ feature ensures a language specific 'cut-off' point for upwards closure.[30]

[30]In Dali and Mathieu (2021b) the features for the plural and plural of the singulative have $[\![+atomic]\!]$ rather than $[\![-atomic]\!]$. I believe this to be a typing error, and have corrected it in this thesis.

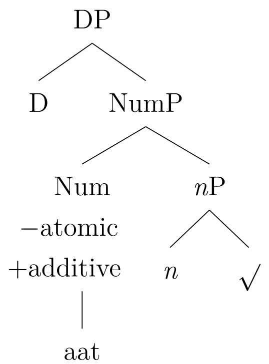


Figure 4.12: Dali and Mathieu's structure for plural terms

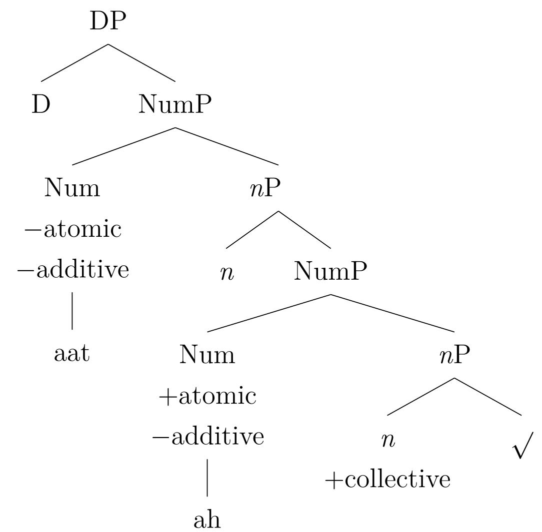


Figure 4.13: Dali and Mathieu's structure for the plural of the singulative

Dali and Mathieu (2021b)'s approach to collective/singulative structures is an improvement to the account found in Mathieu (2012a, 2013, 2014), as singulative marking will only ever apply to the 'correct' root *i.e.* via allosemic insertion in the case of a $num_{[\![+atomic]\!]}$ projection above $\boldsymbol{n}_{[\![+collective]\!]}$. That is, *e.g.* 'ASAD-AH could never arise as the associated root concept does not combine with a $\boldsymbol{n}_{[\![+collective]\!]}$ head, but simply a $\boldsymbol{n}$, and so structures built from ASAD$^{\sqrt{c}}$ ('LION$^{\sqrt{c}}$') are not subject to 'singulative marking' as per the allosemic rule in (47). While this at first seems to be a simple and elegant solution, there there is a concern on the nature of the application of $\boldsymbol{n}$ and $\boldsymbol{n}_{[\![+collective]\!]}$. There is no indication in Dali and Mathieu's account of how or which roots may combine with $\boldsymbol{n}_{[\![+collective]\!]}$ in the first instance. What stops $\boldsymbol{n}$ rather than $\boldsymbol{n}_{[\![+collective]\!]}$ from applying to roots that we would ultimately want to become collectives, and vice versa? Explicitly, why can ASAD$^{\sqrt{c}}$ not combine with $\boldsymbol{n}_{[\![+collective]\!]}$, and why should BAṬ$^{\sqrt{c}}$ ('DUCK$^{\sqrt{c}}$') combine with $\boldsymbol{n}_{[\![+collective]\!]}$? This issue is crosslinguistically problematic, too. As shown in Table 4.6 and Table 4.8 in §4.2, the distribution of the collective/singulative system in Welsh and Arabic is not uniform. Under Dali and Mathieu (2021b)'s approach, Welsh MOCH$^{\sqrt{c}}$ ('PIG$^{\sqrt{c}}$')

combines with $\boldsymbol{n}_{[\![+collective]\!]}$, but Arabic KHANZIIR$^{\sqrt{c}}$ ('PIG$^{\sqrt{c}}$') does not. There is no clear reason why the application of $\boldsymbol{n}_{[\![+collective]\!]}$ should differ crosslinguistically.

Further, it is unclear how a recursive [$\boldsymbol{n}$ +***num***] simplifies matters. First, this solution flouts Economy. The projection of a recursive $\boldsymbol{n}$ in essence 'undos' or 'resets' the semantics of the first ***num*** projection. In theory, paucal semantics should be able to apply in the first projection of ***num***, but in the case of the plural of the singulative they clearly do not. The only reason a recursive $\boldsymbol{n}$ seems to be necessary it to correctly predict singulative marking appearing within the plural of the singulative. Second, while Dali and Mathieu's focus is undoubtedly (Tunisian) Arabic, the authors briefly mention that Welsh has parallel structures for its singulatives, collectives and the plural of the singulative. Yet, as shown in §2.3.2, the Welsh individuating singulative does not pluralise. While we could posit that Arabic has a recursive $\boldsymbol{n}$, but Welsh does not, we cannot easily say why this should be so. There is no clear answer to this situation to be gleaned from Dali and Mathieu's proposal. Therefore, this account ultimately faces the same challenges as Mathieu (2013, 2014). That is, while the tools are there to derive rich crosslinguistic variation, the tools are too powerful and it is not clear how or when these tools should be restricted.

4.3.3 Constructionism summary

The appeal of constructionist accounts lie their ability to capture crosslinguistic variation in such a way that there are minimal mechanisms. There is no reliance on construed ontologies, nor is there a secondary analysis for the behaviour of generalised classifier languages. Rather, all languages are assumed to have the same 'source' for count and mass terms: an under specified lexical root that interacts with a limited number of functional heads. By this measure, constructionist accounts are successful in capturing the fact that object mass, collective, and general number terms share primitive features of number neutrality and uncountability by conflating their syntactic structure (*e.g.* a structure which lacks ***div*** of ***num*** heads). Similarly, the likeness of singulative markers and classifiers are captured as 'divisional' functional heads (*e.g.* ***div*** heads or ***num*** realisations).

Despite the predictive power of constructionist approaches, the accounts presented do not capture the manners in which construed ontological fact influences nominal number marking, such as construed aggregation. Even if the singulative marking is captured by a functional projection of ***div***/***num***, the question of *why* certain terms are marked in the singular and others are not still arises. It is not enough to posit contrasting formal features $\boldsymbol{n}/\boldsymbol{n}_{[\![+collective]\!]}$ and allosemic rules, as the distribution of $\boldsymbol{n}$ and $\boldsymbol{n}_{[\![+collective]\!]}$ needs to be accounted for. A further problem with the constructionist account is there is no way to capture the difference between stuff mass and object mass/collective reference: they are both categorised by the lack of divisional projection. It would then seem that the complete lack of construed ontological influence inhibits constructionist approaches.

4.4 Going forward: what's desired?

Two approaches to collective/singulative systems have been presented in the context of discussion of the lexicalist and constructionist approaches to nominal number. The lexicalist approach assumes with Quine (1960) that number features are 'built in' at the lexical level. As such, any generative component must determine countability status at a pre-syntactic level. Conversely, the constructionist approach implicitly disagrees with Quine by taking all number features as compositionally derived during syntactic computation.

The constructionist and lexicalist approaches to nominal number are complementary in their strengths and weaknesses: lexicalist approaches shine in ability to capture perceptual/construed reality and cultural influence on number marking, while constructionist accounts shine in their ability to seriously reduce the mechanisms that allow for crosslinguistic variation. These strengths and weaknesses are compounded when discussing the collective/singulative systems of Welsh and Hejazi/Modern Standard Arabic. Grimm (2012b)'s lexicalist approach captures the necessity of construed aggregation as a component in collective/singulative 'coding'. It is, however, unclear what the limits are for this account with crosslinguistic variation. The aggregation construal for singulative languages and number marking languages is in danger of being unfalsifiable, and a separate system needs to be defined for generalised classifier languages in which construed aggregation is irrelevant to nominal number, essentially not capturing the ways in which object mass/collective systems are similar to general number systems. For constructionist accounts, the issues are flipped. Mathieu (2012a, 2013, 2014)'s account captures that the object mass, collective and general number terms are subsumed under one semantic category (number neutral uncountable) which is realised as a single structure crosslinguistically, all while unifying the function of the singulative and generalised classifiers under the same syntactic/semantic operation. Presumably the same strengths hold for Dali and Mathieu (2021a)'s account (Dali and Mathieu do not discuss generalised classifier languages). Nonetheless, neither account can successfully predict how or why only certain terms exhibit collective/singulative marking in the first place. What is then the best way to move forward? I suggest the minimal requirements for theory of number are as follows:

Requirement for a system of number:

R1 As per §4.1.2, the substance/object distinction is pre-linguistic, but is nonetheless relevant to language.

R2 As per §2.2.5, at their most primitive level: collectives, object mass, and general number are all number neutral uncountable terms; singulative markers and general classifiers share the same minimal features/functions.

R2b As per §2.2.4, number neutral uncountable terms have access to the (pre-theoretical) objects they denote.

R2c There is blanket number neutral uncountable assignment for nominal terms terms in some languages (*e.g.* generalised classifier languages).

R2d For some languages, construals of aggregation are necessary for number neutral uncountable class (*e.g.* singulative and number marking languages). As per Grimm (2012b, 2018), languages which are sensitive to construed aggregation may be sensitive to different levels of construed aggregation.

By the requirements in R1 and R2, it is not immediately clear whether the lexicalist or constructionist account is more suited to a theory of number. At first, the lexicalist approach seems to win outright. Clearly, the incorporation of construed ontologies into the linguistic system is necessary for [R2b, R2d]. Yet, a constructionist approach, where there are minimal syntactic and semantic primitives to derive number features would be desirable to capture the overarching primitive assumption of [R2].

It would seem, then, that the basic assumptions of a lexico-meaning approach and a constructionist are complementary in their strengths and weaknesses. A non-trivial pursuit of nominal number would be to incorporate the strengths of both approaches into a fusion approach which adopts the conceptual principles that underlie lexico-meaning approaches to number while also benefiting from the crosslinguistic flexibility of constructionist approaches.

4.5 Summary and next steps

This chapter explored consequences and applications of Quine (1960)'s two original conjectures. The first of Quine's conjectures, that children learn the substance/object distinction through the acquisition of language was shown to be untenable. The discussion concluded, in line with standard psychological and cognitive study, that the substance/object distinction is a pre-linguistic notion.

Quine's second conjecture, that count terms come with individuation 'built in' and mass terms do not was discussed from the point of view of the of the lexicalist and constructionist approach to nominal number, with a particular focus on collective/singulative languages. It has been shown that prominent lexico-meaning and constructionist approaches to singulative languages are problematic in complementary ways. Grimm (2012b)'s lexico-meaning approach captures conceptual facts, such as construed aggregation being an important aspect of the collective/singulative class, but leaves little in the discussion of crosslinguistic predictions of exceptions and how the collective/singulative class relates to general number in generalised classifier languages like Mandarin Chinese. Alternatively, constructionist accounts have more powerful tools to deal with crosslinguistic variation, but have little scope to incorporate construed ontological fact. The final proposal of this chapter is that as the strengths of the lexico-meaning approach and constructionist approach are complementary, a view to a fusion account is desirable.

Chapter 5

Building a system: Subpredicative Composition

This chapter offers an enriched version of de Vries and Tsoulas (2021)'s `SUBPREDICATIVE ICEBERG SEMANTICS` approach, which is turn is a constructionist adaption of Landman (2011, 2016, 2020, 2021)'s lexicalist `ICEBERG SEMANTICS` model for the structure of stuff mass, number neutral uncountable and count terms in and across languages. In this way, the account proposed intends to be a hybrid model that synthesises the crosslinguistic power of the constructionist approach while also incorporating the effects of perceptual/construed reality in line with the intuitions of the lexicalist approach.

The development of Subpredicative Iceberg Semantics in this chapter in inspired by the fact that construals of aggregation are variably relevant to the number neutral uncountable class crosslinguistically. In singulative languages, construals of aggregation are highly relevant, where terms for collective aggregates and granular aggregates are typically fall within the number neutral uncountable category. In number marking languages, construals of aggregation are relevant in so much that terms for granular aggregates typically fall within the number neutral uncountable class. In generalised classifier languages, construals of aggregation are irrelevant to the number neutral uncountable class, as all (object denoting) terms fall within the number neutral uncountable class. This observational behaviour indicates that construals of aggregation and their effects on crosslinguistic number are to some degree on a scale. Taking inspiration from Grimm (2012b), this chapter incorporates the idea of scaled construed aggregation into the system of nominal number.

In §5.1, I review the framework in which I will eventually house a theory of nominal number. Both Landman (2020)'s Iceberg Semantics and de Vries and Tsoulas (2021)'s Subpredicative Iceberg Semantics adaptation are presented. The Subpredicative Iceberg Semantics system is shown to be a powerful tool by combining the construed ontological intuitions of lexico-meaning approaches and the predictive power of constructionist approaches. Important to the analysis put forward in this chapter is that in Subpredicative Iceberg Semantics the locus of crosslinguistic variation regarding the count/object mass distinction is identified as a functional projection, **f:set*, in subpredicative nominal composition.

In §5.2, I propose advancements to Subpredicative Iceberg Semantics by manipulation of the semantics of **f:set*. I suggest that the content of **f:set* varies crosslinguistically. In generalised classifier languages **f:set* always projects, accounting for all object denoting terms falling within the number neutral uncountable class. In number marking and singulative languages, **f:set* is associated with a set of presuppositions such that it may only project for a restricted set of nominal derivations. The primary difference between number marking languages and singulative languages are the exact presuppositions associated with **f:set*: number marking languages have a presupposition of granular aggregation, while singulative languages have a presupposition of granular or collective aggregation.

In §5.2.2, I incorporate the fact that while construals of aggregation are necessary for number neutral uncountable derivation in number marking and singulative languages, it is not sufficient as not all construed aggregate individuals have number neutral uncountable representations in singulative languages. To account for exceptions, I propose a conglomerate measure of individuation, MOI(P), which is calculated from sub-measures of animacy, size, and interaction. If individuals with respect to a given property have a high measure of individuation, then number neutral uncountable derivation fails. While MOI(P) is based on speaker construals, it is suggested that it is primarily a linguistic (and not perceptual) phenomenon as the precise recipe of MOI(P) calculation is language, and not culture, dependent.

In §5.3, I discuss the differences between the Subpredicative Iceberg Semantics approach presented here and Grimm (2012b)'s approach, concluding that the Subpredicative Iceberg Semantics approach affords a flexibility to the exceptions of the collective/singulative class that are otherwise hard to explain. The discussion also briefly touches upon Gathercole et al. (2000)'s experiment which shows that Welsh speaking children have tacit knowledge that their language is sensitive to construals of aggregation.

5.1 Setting the Scene

5.1.1 A hybrid approach?

As discussed in chapter 4, formal approaches to nominal number can be divided into two broad types: the lexicalist approach, where number features are lexically given (Bloomfield, 1984; Chierchia, 1998a,b, 2010b, 2017, 2021; Khrizman et al., 2015; Landman, 2011, 2016, 2020, 2021; Grimm, 2012b,a, 2018; Rothstein, 2010, 2011, 2017, 2021, *a.o.*), and the constructionist approach, where all number features are derived compositionally (Bale and Barner, 2009; Borer, 2005a; Borer and Ouwayda, 2021; Dali, 2020; Mathieu, 2012a; Martí, 2020; Pelletier, 2012, *a.o.*). The strengths and weaknesses of the lexicalist and the constructionist approach are complementary. Lexicalist approaches shine in their ability to map construed ontologies and linguistic ontologies, but struggle with accounting for count/mass flexibility both in and across languages. Constructionist approaches on the other hand have far greater tools to account for crosslinguistic variation, but struggle to capture conceptual factors.

To account for the strengths and shortcomings of both lexicalist and constructionist approaches, de Vries and Tsoulas (2021) offer a synthesis of the strengths of both approaches; de Vries and Tsoulas take a Borer (2005a) inspired constructionist framework but fuse it with a Landman (2011, 2016, 2020, 2021) lexicalist style semantics, which de Vries and Tsoulas call `SUBPREDICATIVE ICEBERG SEMANTICS`[1]. The modifications in Subpredicative Iceberg Semantics allow for the deriving of conceptual factors such as Chierchia (2010b)'s mapping property (where all substances are represented by mass terms) though minimal syntactic/semantic primitives. de Vries and Tsoulas's account thus captures the desirable conceptual facts in such a manner that uses very few generative tools.

5.1.2 Iceberg Semantics

The formalism that underlies de Vries and Tsoulas (2021)'s Subpredicative Iceberg Semantics is based on Landman (2011, 2016, 2020, 2021)'s `ICEBERG SEMANTICS`. A descendant of Linkian tradition, Landman's Iceberg Semantics is both lexicalist and Boolean in nature, where the domain of entities, B, is a complete Boolean algebra ordered by the part-of relation, $\sqsubseteq$, which satisfies the axioms of being reflexive, antisymmetric, transitive, and unique.[2] In Iceberg Semantics, Landman distinguishes between the set of objects of B, and the set of elements of B. The set of objects is a sub-semi-lattice of B, defined as B minus the bottom element in (1).

(1) Let B be a Boolean algebra and $X \subseteq B$.
➢ X^+, the set of *objects* in X, is given by: $X^+ = X - 0$
From Landman (2020, p.18)

The denotation of a nominal term is not some $X \subseteq B$ as in the so-called `MOUNTAIN SEMANTICS` of Link (1983); Krifka (1995) and Chierchia (1998a, 2010b) *a.o.*. Rather,

[1] In de Vries and Tsoulas (2021), the account is referred to as `EXOSKELETAL ICEBERG SEMANTICS`. However, from discussions with George Tsoulas (*p.c*), the term `SUBPREDICATIVE ICEBERG SEMANTICS` is now preferred.

[2] See §3.2 for a more thorough overview of the basics of Boolean models and §4.1.3 for mereological methods in linguistics a la Link (1983).

the denotation of a nominal term is of an `ICEBERG SET` (i-set), which is a pair X of **body** and **base**, formalised as ⟨body, base⟩. Both the **body** and the **base** are sets of mereological sums which are subsets of B, and the **body** generates the **base** under sum. The formal definition of i-set is given in (2), and closure under sum is defined in the normal Linkian way as in (3).[3]

(2) Let B be a complete Boolean algebra.
➢ An i-set is a pair $X = \langle \mathbf{body(X), base(X)} \rangle$ where:
$\mathbf{body(X)} \subseteq B$ and $\mathbf{base(X)} \subseteq B$ and $\mathbf{body(X)} \subseteq {}^*\mathbf{base}(X)$
and $\sqcup\mathbf{body(X)} = \sqcup\mathbf{base(X)}$.
From Landman (2020, p.143)

(3) `ALGEBRAIC CLOSURE`
$^*P \overset{\text{def}}{=} \{x\!: \exists P'\ P' \neq \varnothing \wedge P' \subseteq P \wedge x = \sqcup\, P'\}$
The set that contains all sums of things in P
From Champollion and Krifka (2016), adapted

In Iceberg Semantics, all (non-)countability features are defined in terms of disjointedness distinctions made in either the **body** or the **base** of an i-set. In a nutshell, only if the **body** of an i-set is disjoint, then the interpretation will be singular; if the **body** of an i-set is not disjoint it is interpreted as non-singular (4)[4]. On the other hand, countability distinctions are made in the **base**, where if the **base** of an i-set is disjoint, then that predicate is count, mass otherwise (5).

[3]For full transparency, Landman (2020) defines closure under sum as in (i), with reference to the domain of discourse, B. The difference between Landman's definition of sum and sum as defined in (3) is minimal and will not bother us here.

(i) closure of sum of X, *X, is given by:
$^*X = \{b \in B : \text{for some } Y \subseteq X : b = \bigsqcup Y\}$
From Landman (2020, p.21)

[4]Landman (2020) does not actually give a general definition for singular and non-singular semantics. I supply the definition given in (4) based on the definition in (5) and information Landman (2020, p.143) gives for the semantics of plurals.

(4) Let X be an i-set
➢ X is *singular* iff **body**(X) is disjoint
➢ X is *non-singular* iff X is non-null then X is not singular.

(5) Let X be an i-set
➢ X is *count* iff **base**(X) is disjoint
➢ X is *mass* iff X is non-null then X is not count.
From Landman (2020, p.165, adapted)

The motivation for assigning different features of countability to the **body** and the **base** comes down to the need for the semantics to explicitly keep track of the distribution set, *i.e.* the **base**, so that the terms in which the **body** is counted is fully available at any given point. In other words, the **base** needs to supply, in context, what counts as a maximally disjoint individual so that counting is licensed.

The status of the **body** and the **base** derive the countability properties of singular, countable plural, number neutral uncountable and stuff mass terms. To see how this works, consider first singular count CAT and countable plural CATS in English. As countable terms, CAT and CATS are i-sets with a disjoint **base**.[5] Singular count and plural count terms differ only in the disjointedness status of the **body**, where *e.g.* singular CAT has a disjoint **body** (refers to singularities) and plural term CATS has an overlapping **body** (closed under *).[6] The formal notation for English CAT and CATS is given in (6) and (7).

[5]That singular terms are restricted to combine with numeral ONE is due to the nature of the **body**. For singular count terms, the **body** refers to singularities, and so is incompatible with combining with higher numerals.

[6]This closure under sum is inclusive (contains reference to the set of singularities). Presumably the exclusive interpretations of plurals in Iceberg Semantics are derived by some (scalar) implicature. Landman (2020) does not discuss how exclusive interpretations are derived.

(6) Singular count nouns, *e.g.* CAT =
⟨CAT, CAT⟩

(7) Existential plural count nouns, *e.g.* CATS =
⟨*CAT, CAT⟩

In a toy universe with only three cats, a = Alice, b = Bella and c = Carlisle, the **body** and the **base** of singular and plural terms are expressed in (8), with visualisation shown via Hasse lattice in Figure 5.1 and Figure 5.2 where the outer selection represents the **body** and the inner selection represents the **base**.

(8) In a toy universe where there are only three cats

Singular count CAT

⟨CAT, CAT⟩

body = {a, b, c}

base = {a, b, c}

Plural count CATS

⟨*CAT, CAT⟩

body = *{a, b, c} = {a, b, c, $a \sqcup b$, $a \sqcup c$, $b \sqcup c$, $a \sqcup b \sqcup c$}

base = {a, b, c}

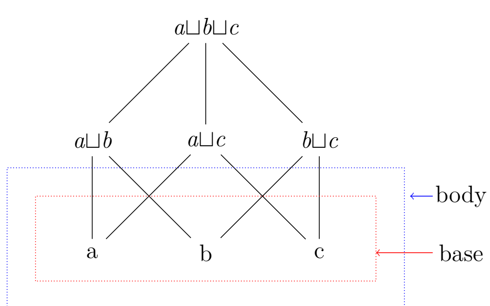


Figure 5.1: Landman's iceberg structure for singular count terms

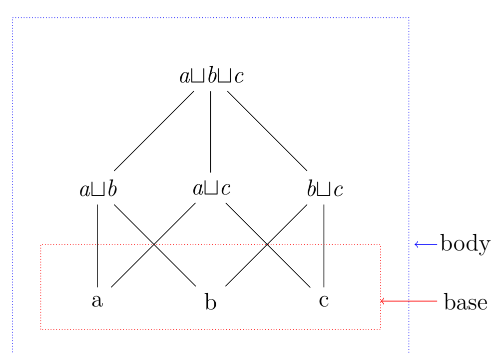


Figure 5.2: Landman's iceberg structure for plural count terms

Mass terms are defined by the nature of the **base**, which is only quasi-influenced by the structure of matter. The **base** corresponds to a `PERSPECTIVE` upon construed ontological structure of matter. To see how this works, first consider a stuff mass term like WATER. For Landman (2020), stuff mass terms arise because the substances they denote are perceived as overlapping at every context, *i.e.* there is no good construed standard for what counts as a disjoint instance of water that holds across contexts. Consequently, the **base** is necessarily overlapping, leading to stuff mass interpretation for WATER.[7] In this way, Landman derives the Iceberg Semantics version of the mapping property (see §4.2.1): individuals that are not disjoint at every construed ontological context must always come out as mass.

(9) Substance Mass, *e.g.* WATER =
⟨WATER, WATER⟩, where the **base** is not disjoint at every context

For object mass terms, the picture is different. There are at least two types of object mass: `GROUP NEUTRAL` and `SUM NEUTRAL`. For group neutral terms, Landman (2020, p.191) states there may or may not be vertical overlap in the **base** in any given context. Examples given for group neutral terms include KITCHENWARE, which is said to have ambiguity as what counts as an individual due to *e.g.a-pan*$\sqcup$*pan-lid* counting as an individual in one context, but two individuals in another. In this way, group neutral nouns are mass due to the same logic that substance denoting terms are mass: the individuals are not disjoint at every context, and so the **base** is overlapping leading to mass coding.

(10) Group Neutral, *e.g.* POTTERY, KITCHENWARE =
⟨*POTTERY, POTTERY⟩, where the **base** is not disjoint at every context

For sum neutral terms, Landman (2020, p.194) posits that there is no construed overlap of individuals across contexts that leads to the **base** overlapping, but rather the perspective between singular and plural is made irrelevant by a grammatical

[7]As the **body** is then generated from the **base**, this necessarily leads to number neutrality for stuff mass terms via the definition of the i-set generation under sum in (2).

reflex that lexically closes the **base** under sum. The existence of sum neutrality in Landman's account is the locus of the required arbitrariness in the system where disjoint entities at the construed ontological level may nonetheless be represented by object mass terms at the linguistic ontological level. The example that Landman provides for English is POULTRY. The idea is that while farm birds have a disjoint standard across contexts, the **base** of the resulting i-set is nonetheless arbitrarily closed under sum. Landman offers no discussion as to why sum neutral terms exist, nor is there any prediction as to which terms may become sum neutral. Presumably, Landman treats all object denoting terms in Mandarin Chinese (which are not vertically overlapping across contexts) as sum neutral.

(11) Sum Neutral, *e.g.* POULTRY =
$\langle$*FARM BIRD, *FARM BIRD$\rangle$, where **base** is closed under *

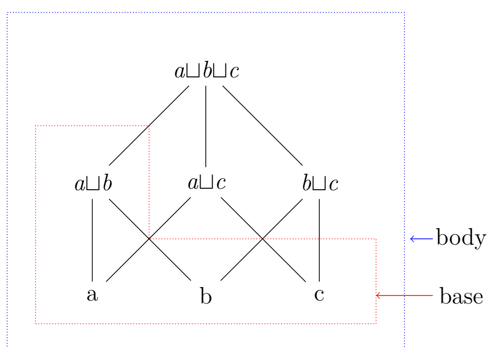


Figure 5.3: Landman's iceberg structure for group neutral terms

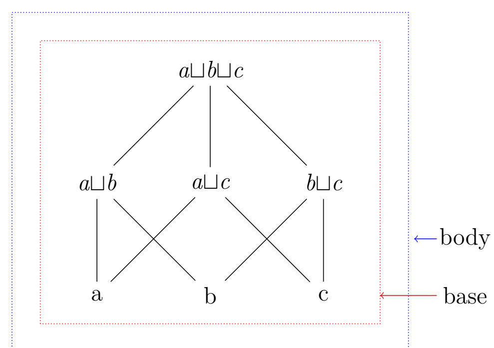


Figure 5.4: Landman's iceberg structure for sum neutral terms

5.1.3 Subpredicative Iceberg Semantics

Iceberg Semantics is a powerful tool for describing predicates across two dimensions that are loosely independent: countability (via distinction in the **body**), and number neutral/singular reference (via distinctions in the **base**). Iceberg Semantics, however, faces the same criticisms typical of lexicalist accounts: it struggles to predict crosslinguistic variation. For example, KITCHENWARE is object mass in English but German KÜCHENGERÄT is singular count. It is unclear how such variation can be accounted for. Should we say that for English speakers kitchenware items are not disjoint at every context, but for German speakers they are disjoint at every context? Surely not.[8]

Though the crosslinguistic aspects of Iceberg Semantics are suspect, the formal mechanisms are incredibly powerful tools. It turns out that even though Iceberg Semantics is not a feature-based account, it is sympathetic to feature-based approaches, as it essentially offers $[\![\pm countable]\!]$ and $[\![\pm number\ neutral]\!]$ distinctions. Further, as the **body** is generated from the **base**, Iceberg Semantics is hierarchically structured, and therefore even though Landman (2020)'s Iceberg Semantics is lexicalist in nature, it is in fact adaptable to a constructionist model where i-sets are compositionally derived, rather than lexically given. This is exactly the route that de Vries and Tsoulas (2021)'s take with their `SUBPREDICATIVE ICEBERG SEMANTICS` model for countability.

de Vries and Tsoulas (2021)'s account account is `SUBPREDICATIVE`, as it concerns the semantics/syntax interface below the structure of a predicate. The subpredicative nature is however described in the constructionist fashion. As Borer (2005a) before them, de Vries and Tsoulas take all lexical items as starting their derivational lives as classless roots, *i.e.* they are root concepts. Unlike Borer's account, these root concepts are not taken to be undifferentiated 'stuff'. Rather, a root concept is modelled as a union of construed ontologies encompassing all associations of a particular concept. For example, the root concept CAT$^{\sqrt{c}}$ will be associated with all physical manifestations of cats and cat-events, if they exist; the root COMPUTER$^{\sqrt{c}}$ will be associated with physical manifestations of computers, and computing events, if they exist; the root PEN$^{\sqrt{c}}$ will be associated with physical pens, and pen-events, if they exist, so on and so forth *ad nauseam*.

[8]See §4.2.2 for more discussion on the issues of the construed disjointedness influence on nominal number.

Roots are turned into grammatical objects by a `FILTER OPERATION`. With respect to the physical domain there are two filters proposed: the IND filter (12) and the SUB filter (13). Assuming that INDIVIDUAL(x) and SUBSTANCE(x) are are sets of pre/extra-linguistic objects and substances respectively, the IND filter applies to root concepts to reveal only *individuals* relevant to the root, and the SUB filter applies to root concepts to reveal only *substances* relevant to the root. The output of the filter operation is a `FILTERED SET`.[9]

(12) $\textsc{ind} \stackrel{\text{def}}{=} \lambda C \lambda x[C(x) \wedge \textsc{individual}(\mathrm{x})]$
From de Vries and Tsoulas (2021, p.13)

(13) $\textsc{sub} \stackrel{\text{def}}{=} \lambda C \lambda x[C(x) \wedge \textsc{substance}(\mathrm{x})]$
From de Vries and Tsoulas (2021, p.13)

Filtered sets are not linguistic items as such, but are the 'building blocks' of predicates. To turn filtered sets into linguistic items, de Vries and Tsoulas propose the `PREDICATE FORMATION FUNCTION`, PRED_FORM, as defined in (14), which creates an i-set in the Landman manner.

(14) Let F be a filtered set
PRED_FORM(F) = $\langle$F, F$\rangle$
From de Vries and Tsoulas (2021, p.13), adapted

During the composition of stuff mass terms, the SUB filter applies to a root concept, and the output is a filtered set of non-disjoint (overlapping) individuals. The non-disjointedness of the filtered set is preserved in the **base** during predicate formation, necessarily leading to a substance mass i-set. An example of stuff mass composition for WATER is presented in Figure 5.5.

[9] de Vries and Tsoulas (2021) do not define the pre/extra linguistic set of substances/individuals. I assume a mereotopological definition of substances and individuals as discussed in §3.5, and therefore take both the set of pre/extra linguistic individuals and the set of pre/extra linguistic substances to be also defined in mereotopological terms.

During the composition of singular count terms, the IND filter applies to a root concept, and the output is a filtered set of disjoint individuals. The disjointedness of the filtered set is preserved in the **base** during predicate formation, leading to a singular count i-set. An example of singular count composition for CAT is presented in Figure 5.6.

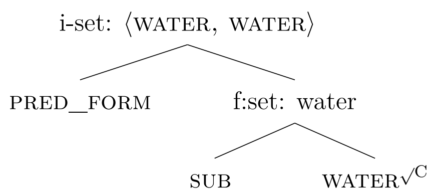


Figure 5.5: de Vries and Tsoulas' structure for stuff mass terms

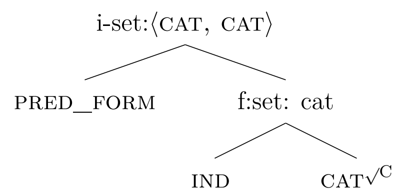


Figure 5.6: de Vries and Tsoulas' structure for singular count terms

de Vries and Tsoulas posit that the initial composition of number neutral uncountable (*i.e.* object mass and general number) and number neutral countable (*i.e.* plural) terms mirrors singular count terms: the IND filter applies to a root concept, and the output is a filtered set of disjoint individuals. The composition then differs in the application of Linkian sum closure, *, which may take place immediately before or immediately after PRED_FORM.

In the case of number neutral uncountable terms, * occurs to the filtered set, closing it under sum, as **f:set* in (15). This results in an overlapping set (singularities and their joins are in the set). The overlapping nature of the output of **f:set* is preserved as the **body** and **base** during predicate formation via PRED_FORM. The overlapping **body** ensure number neutrality/plurality, and the overlapping **base** ensures uncountability. An example of number neutral uncountable composition for is presented in Figure 5.7.

In the case of number neutral countable terms, pluralisation occurs after predicate formation to the **body** only, as schematised in (16).[10] The overlapping **body** ensures

[10]The semantics of *BODY here are schematic, and are developed further in chapter 6.

neutrality/plurality. Crucially, as only the **body** is closed under sum, the **base** remains a disjoint set of individuals. This disjoint **base** ensures countability. An example of number neutral countable composition is presented in Figure 5.8.

(15) Let P be a filtered set,
**f:set* = *P

(16) *BODY = ⟨**body**, **base**⟩ → ⟨***body**, **base**⟩

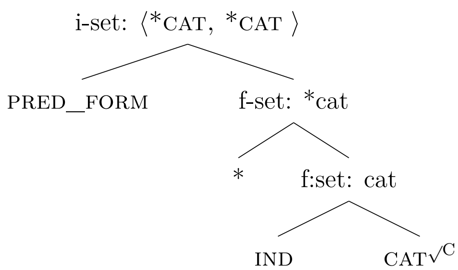


Figure 5.7: de Vries and Tsoulas' structure for number neutral uncountable terms

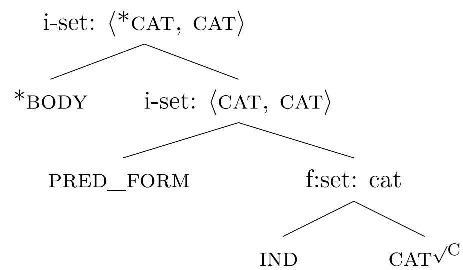


Figure 5.8: de Vries and Tsoulas' structure for number neutral countable terms

de Vries and Tsoulas (2021)'s Subpredicative Iceberg Semantics system combines the strengths of lexicalist and constructionist approaches. Subpredicative Iceberg Semantics allows for the crosslinguistic flexibility of constructionist accounts. For example, all number neutral uncountable terms, *e.g.* English object mass and Mandarin Chinese general number, are analysed as i-sets which have a **f:set* projection. de Vries and Tsoulas take the stance that all object denoting predicates in Mandarin Chinese have a compulsory covert **f:set* projection. In this way, number neutral uncountable terms are all derived in the same manner, and the locus of linguistic variation between count and mass terms is at the level of **f:set*.

At the same time, Subpredicative Iceberg Semantics enjoys benefits typically not associated with constructionist accounts. Firstly, Subpredicative Iceberg Semantics does not equate mass terms with the lack of projection as in Borer (2005a,b, *et. seq.*), but rather with an [IND + *f:set*] complex in the case of number neutral uncountable and a [SUB] projection in the case of canonical stuff mass. The consequence of this is that as number neutral uncountable and stuff mass terms are structurally and semantically distinct, and so we expect differences in their uses and interpretations. For example, any argument that number neutral uncountable terms should not have access to objects and must denote 'stuff' due to conflation with stuff mass structure is obliterated. Instead, as number neutral uncountable terms necessarily have an IND projection then they always have prelinguistic objects/individuals defined. Under this approach, `STUBS` and comparison constructions can straightforwardly be analysed as applying to terms which have an the IND filter projection (which includes singular count, plural count alongside number neutral uncountable terms).

The Subpredicative Iceberg Semantics model also predicts the mapping property. This is because while in theory any filter can apply to any root, construed ontological considerations will necessitate that some root/filter combinations are incompatible. For example, a substance denoting root like WATER$^{\sqrt{c}}$ may have no pre-existing individuals associated with it, and so may never compose with the IND filter (or if it does, it comes out empty). From this reasoning, all substance terms, in all languages, are predicted to combine only with the SUB filter, and therefore come out as stuff mass.

5.2 Extending Subpredicative Iceberg Semantics

At their most basic level, collective terms in Welsh and Arabic are number neutral uncountable terms (see chapter 2). A natural extension to Subpredicative Iceberg Semantics is to incorporate collective terms as i-sets of type ⟨***body**, ***base**⟩, where the overlapping **body** ensures number neutrality and the overlapping **base** ensures uncountability. An example derivation of Welsh collective HWYAID (duck.COL = *'duck(s)'*) is given in Figure 5.9.

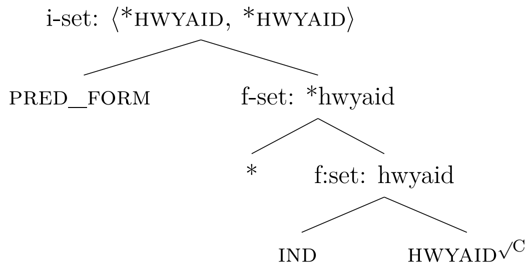


Figure 5.9: Collective terms in Subpredicative Iceberg Semantics

This immediately raises a question: how does the syntax know when to project **f:set*? Similarly, why should **f:set* apply to the filtered set over DUCK$^{\sqrt{c}}$ in Welsh and Arabic, but not English? I suggest that the simplest solution follows the wisdom found in Grimm (2012b), *i.e.* that various types of construed aggregation are highly relevant to nominal systems. In what follows, I incorporate perceptual aggregation into the model such that construals of aggregation are relevant to **f:set* projection.

5.2.1 The basic architecture for collective terms

For de Vries and Tsoulas, the projection of **f:set* is compulsory for generalised classifier languages like Mandarin Chinese, but its projection is not compulsory in number marking languages like English. The picture must be more nuanced than this. Some languages have very few or no number neutral uncountable terms such as Greek (see Alexiadou, 2015; Chierchia, 2010b, 2021; Erbach, 2021; Tsoulas, 2009, for discussion) or presumably Yudja (Lima, 2010, 2014). For these languages, **f:set* must be highly restricted, or even not projected at all. On the other hand, the collective/singulative class in Welsh and Hejazi/Modern Standard Arabic is a highly productive number neutral uncountable nominal system. The appearance of number neutral uncountable terms therefore appears to be scaled across language, as modelled in Figure 5.10. Note, Figure 5.10 is not meant to imply equidistance between some measurement of productivity of **f:set* between *e.g.* English and Arabic on the one hand and Arabic and Mandarin Chinese on the other hand. Nor is it intended to imply that Arabic and Welsh are identical in their productivity of the

**f:set*. Rather, the scale in Figure 5.10 is schematic of where certain languages compare to each other in their **f:set* projection productivity.

Figure 5.10: The scale of **f:set* projection across languages

The scaled nature of the appearance of number neutral uncountable terms across languages is captured by the the prevalence of **f:set* projection. Following de Vries and Tsoulas, I take **f:set* in generalised classifier languages to be free, and therefore it always projects. I refer to these types of language as `FREE`* languages. In the case of number marking languages (English), singulative languages (Welsh, Arabic) I submit that **f:set* projects to a varying degree. I refer to these types of languages as `PRESUP`* languages. To ensure that **f:set* only projects in some cases in `PRESUP`* languages, the projection of **f:set* must come with a set of (language specific) presuppositions that restrict its legal projections.

The semantics for **f:set* in `FREE`* languages is presented in (17). The set of presuppositions associated with **f:set* in `PRESUP`* languages, however, is modelled as a series of disjunctions as schematised in (18). In `PRESUP`* languages, there is variability in the number of disjunctions associated with **f:set*. Number marking and singulative languages (English, Welsh, Arabic) have a varying number of presuppositions associated with **f:set* such that **f:set* projects in more cases in singulative languages than number marking languages. Trivially, as `FREE`* languages always project **f:set*, the semantics given for the `FREE`* version of **f:set* in (17) may be lowered into a series of infinite conjunctions. In no-object-mass languages such as Greek (and presumably Yudja), where there are very few to no object mass terms, **f:set* must be either not parametrised, or it is parametrised and must have the most restrictive disjunctions associated with it.

(17) `FREE`* languages (Mandarin Chinese):

$$\textit{*f:set} \rightarrow \lambda \mathrm{P}\begin{cases} \lambda x[{}^{*}P(x)] \end{cases}$$

(18) `PRESUP`* languages (English, Welsh, Arabic):

$$\textit{*f:set} \rightarrow \lambda \mathrm{P}\begin{cases} \lambda x[{}^{*}P(x)] & \text{if } \textsc{presupa} \vee \textsc{presupb} \ldots = \text{TRUE} \\ \bot & \text{otherwise} \end{cases}$$

The task is then to identify the presuppositions associated with **f:set* in and across `PRESUP`* languages. As the distribution of the collective/singulative systems of both Welsh and Hejazi/Modern Standard Arabic typically covers aggregate notional categories such as granular aggregates which typically are in touch (*e.g.* seeds, pulses, grains) and collective aggregates which may not touch but typically come in groups (*e.g.* fruits, vegetables, insects and farmed animals), the immediately obvious integration would be one of construed aggregation. However, the presuppositions associated with **f:set* in `PRESUP`* languages cannot be a simple catch-all aggregation construal, as number marking languages and singulative languages differ on the types of aggregation relevant to number neutral uncountable composition. While Welsh and Hejazi/Modern Standard Arabic have reliable number neutral uncountable coverage of notional collective aggregates, English only has reliable number neutral uncountable coverage for notional granular aggregates, with very few non-granular aggregates entering the number neutral uncountable class.

With this divide, the construed aggregation presuppositions associated with **f:set* must differ across `PRESUP`* languages. To formalise this, consider `THE ORDER OF AGGREGATION` in (19), which is inspired by The Scale of Individuation in Grimm (2012b). The segments of The Order of Aggregation should be seen as the sets of presuppositions languages may associate with **f:set*, and therefore the notional requirements that the individuals denoted by the filtered set must meet to license

**f:set* projection.[11] Taking further inspiration from Grimm (2012b, 2018), The Order of Aggregation is strictly ordered such that if a language recognises collective aggregates as a relevant presupposition of **f:set* projection, then it is the case that the language must also be sensitive to granular aggregates for **f:set* projection. As such, there should be no language which derives notional collective aggregates as number neutral uncountable without also deriving notional granular aggregates as number neutral uncountable. Being strictly ordered, the reverse does not hold: a language may freely associate granular, but not collective, aggregates with **f:set*.

(19) `THE ORDER OF AGGREGATION`
granular aggregates < collective aggregates

In practice, this materialises as the granular aggregate category being a presupposition associated with English **f:set*. This is seen by the tendency for English number neutral uncountable terms to refer to grains (*e.g.* MILLET, BARLEY, QUINOA). The notional collective aggregate category is not associated with English **f:set*, and so fruits, vegetables, insects *etc.* are not typically associated with English number neutral uncountable terms. For Welsh and Hejazi/Modern Standard Arabic, both the granular aggregate and collective aggregate classes are presuppositions associated with **f:set* projection. With this, the minimal differences between number marking languages like English and collective/singulative languages like Welsh and Arabic therefore lies within the presuppositions associated with **f:set*. As these presuppositions are disjunctive, the scaled effect is achieved: the number marking language **f:set* projection criteria entails the singulative language **f:set* projection criteria, which in turn trivially entails the `FREE*` **f:set* projection criteria.

[11]A key difference between The Order of Aggregation and Grimm (2012b)'s The Scale of Individuation is that liquids/substances do not form part of the The Order of Aggregation. This is because by hypothesis, substance denoting terms are derived not via **f:set* projection, but via the SUB filter.

(20) Types of `PRESUP*` languages:

Number marking languages:
Granular aggregate presupposition satisfied

Singulative languages:
(Granular aggregate presupposition) $\vee$ (Collective aggregate presupposition) satisfied

This modification to de Vries and Tsoulas (2021)'s Subpredicative Iceberg Semantics account proposes that uniform mechanisms are available crosslinguistically (*i.e.* the projection of **f:set*), but allows for languages to experience variation on how this mechanism is applied. The crosslinguistic difference in nominal systems is taken to be a difference in *degree* and not *kind*, such that the locus of crosslinguistic variation, *i.e.* **f:set* projection criteria, is scaled in nature.

5.2.1.1 Defining construed aggregation

I now define the construed aggregation presuppositions associated with **f:set* in `PRESUP*` languages.[12] The intuitive difference between notional granular and collective aggregates is that granular aggregates are collections of like-individuations that are typically touching, while collective aggregates are collections of like-individuals which are not typically touching, but are sufficiently near each other. With this, I model the intuitions of `TOUCH` and `NEARNESS` of two individuals, and then incorporate these definitions into one of construed aggregation.

To define intuitive nearness of two entities, I define NEAR.[13] Taking $\mathcal{D}$ to be the disjointedness relation and d to be a standard distance function I model the

[12]This section discusses mereotopological and distance notions. See chapter 3 for background assumptions.

[13]Both Grimm (2012b) and Wągiel (2021b) use the term `PROXIMITY CONNECTED` for parallel relations. As distance is not a connectedness relation in the mereotopological sense (see §3.6 for discussion), I do not use the term `CONNECTED`. Rather, I opt to use the terminology `NEAR RELATION` to emphasise intuitive nearness.

NEAR function as in (21), where two entities are considered NEAR if their distance is less than some amount $\boldsymbol{n}$. I take the value of $\boldsymbol{n}$ to be determined relevant to the context.[14] The point of the flexible value of $\boldsymbol{n}$ is such that the threshold for intuitive nearness differs across contexts. For example, the threshold for intuitive nearness of two ducks is not comparable to the intuitive nearness of two ants. Explicitly, if two ducks are five meters apart, they may be considered near each other, but if two ants are five metres apart, they may not be considered near each other. Further, as the definition for NEAR aims to define whether two distinct individuals are near each other, we should want NEAR to hold only for two distinct individuals, and not to suggest that anything may be NEAR to itself. As such, NEAR specifies that x and y must be disjoint to rule out x and y being the same entity making the value for d = 0, implying that everything is near to itself.

(21) $\textsc{near}(x,y) \stackrel{\text{def}}{=} \mathcal{D}(x,y) \wedge d(x,y) \leq \boldsymbol{n}$
Two disjoint entities, x and y, are near if the distance between them is less than or equal to a contextually salient measure.

To define touch, it may be tempting to follow Grimm (2012b) (and later Wągiel's series of works) and use the mereotopological external connectedness relation in (22), where $\mathcal{C}$ is the mereotopological connectedness relation. According to this definition, two entities are externally connected if they are connected but their interiors do not overlap.

[14]Unlike Wągiel (2021b), I do not call for a predicate in the formulation of (21). This is because it may come in useful to describe a cat and a dog as being near each other, and Wągiel (2021b)'s semantics in (i) could not be expanded to account for such cases, as $\boldsymbol{n}$ is restricted to a single property. Admittedly, such cases are not directly relevant for the composition of number neutral uncountable terms, so nothing hinges on this.

(i) $\textsc{pc}(x, y, P) \stackrel{\text{def}}{=} d(x,y) \leq n(P)$
From Wągiel (2021b)

(22) External Connection
$\textsc{ec}(x,y) \overset{\text{def}}{=} \mathcal{C}(x,y) \wedge \neg\mathcal{C}(\textsc{int}x, \textsc{int}y)$
Two entities x and y are externally connected if they are connected, but their interiors do not overlap.
From Grimm (2012a, p.136)

I will not use external connection to model touching individuals, as it is inappropriate for capturing vanishing closeness. Two grains of rice may be touching to the human eye, with no perceivable space between them, but that does not mean they are connected in the mereotopological sense of the $\mathcal{C}$ relation. To be connected in the mereotopological sense, individuals must either overlap (overlap implies connection) or share a boundary. Placing the issue of open and closed boundaries asides, to be externally connected two individuals must have *no space between them at all*. The definition in (22) is only compatible with entities that have no internal overlap, but cannot be separated. As such, external connection is defined for countries with fiat boundaries (such as France and Spain), but not for two maximally strongly self connected individuals that are touching in an intuitive sense (such as two touching grains of rice in a pile). It is simply not the case that (22) may well describe touching or vanishingly small distances between individuals, such as two touching grains of rice.

Following Casati and Varzi (1999) I take 'vanishingly close' entities to be a matter of metric concern (see §3.6). Therefore, distance functions should be used not only for describing entities that are intuitively near to each other, but also for entities that are vanishingly close, *i.e.* touching.[15] I define CLOSE in (23), which is true in case two individuals have a distance that is approaching 0. In other words, as close as possible for two disjoint individuals to be without imposing absolute topological

[15]There is little formal work on the semantics of touch. Other than Grimm (2012b), there is another formalisation of nearness that relies on distance functions, *i.e.* that of Henderson (2017) who defines a symmetric ADJACENT feature which does not seem to have the issues of external connectedness in the mereotopological sense. Alternatively, there is an account in Bittner (2017) which relies on a supervaluation-based framework of distance and connection.

change which leads to overlap or shared boundaries. As with the NEAR relation, I require the two individuals of the CLOSE relation to be distinct and disjoint.[16]

(23) $\text{CLOSE}(x,y) \stackrel{\text{def}}{=} \mathcal{D}(x,y) \wedge d(x,y) \approx 0$
Two disjoint entities, x and y, are close to each other if the distance between them is approaching 0.

The definitions of CLOSE and NEAR are such that they hold between two distinct individuals, and not of a series of individuals or a notional aggregate. Let us now define what it means to be an aggregate. The definition sought for aggregates should include a.) a shared property, and b.) modelling of the intuition that an aggregate is a collection of disjoint individuals that are typically CLOSE or NEAR to another. Incorporating CLOSE and NEAR into a definition of aggregation should not use universal quantification. Consider the proposed definitions in (24) for granular aggregates and (25) for collective aggregates. Due to the universal quantifier, there is a commitment that every individual with a certain property P is CLOSE/NEAR to another individual of the same property, with no exceptions. This is clearly too strong: a single grain of rice may exist in silo, without touching another grain of rice. Such a rogue grain of rice does not seem to falsify the intuition that rice grains are typically in touch, but it does falsify any claim that *all* grains are in touch with another rice grain. Similar observations hold for ants: while ants typically come in clusters or groups, it would not be unheard of to find a singular ant crawling around the kitchen, separate from his insectile comrades. Again, this counterexample does not falsify the claim that ants typically are close to other ants, only the claim that *all* ants are close to other ants.

[16] The definition in (23) is modelled after an intuition of vanishing closeness where entities are in touch. CLOSE should not be confused with any natural language use of '*close*'. Consider the following scenario. Bella lives in Forks, Washington, USA. Her soulmate Edward is in Volterra, Tuscany, Italy. This is a distance of around 9250km, or so Google tells me. If in the situation that Bella flew into Pisa Airport, she would be a mere 49.9km away from Edward. Upon arriving at Pisa Airport, Bella may well consider herself *close* to Edward in the natural language sense. This interpretation of the predicate CLOSE considers *relative closeness*, which is not intended as part of the CLOSE relation. The definition for CLOSE I give intends to be only vague in the sense of what approximates 0.

(24) $\forall x[P(x) \rightarrow \exists y[P(y) \wedge \textsc{close}(x,y)]$

(25) $\forall x[P(x) \rightarrow \exists y[P(y) \wedge \textsc{near}(x,y)]$

The representation of construals of aggregation are better understood as a cardinality judgement, where the cardinality of the set of like-individuals which are NEAR/CLOSE to another like-individual is greater than the cardinality of the set where like-individuals are not CLOSE/NEAR to another like-individual.[17] The construed aggregation intuition is modelled in (26) for granular aggregates and (27) for collective aggregates. Using real examples, (26) will be true for the property of RICE, CORN, QUINOA *etc.* as individuals that satisfy these properties are CLOSE to like-individuals more often than they are not. Note, the semantics in (26) will only hold for predicates which denote individuals are that typically CLOSE to like-individuals. It will not hold for predicates such as CARROT, ANT, PIG *etc.*. On the other hand, (27) will hold for the CARROT, ANT, PIG properties.

(26) $\textsc{gran}(P) = \text{TRUE iff.}$
$|\{x\colon P(x) \wedge \exists y[P(y) \wedge \textsc{close}(x,y)]\}| > |\{x\colon P(x) \wedge \exists y[P(y) \wedge \neg\textsc{close}(x,y)]\}|$

(27) $\textsc{col}(P) = \text{TRUE iff.}$
$|\{x\colon P(x) \wedge \exists y[P(y) \wedge \textsc{near}(x,y)]\}| > |\{x\colon P(x) \wedge \exists y[P(y) \wedge \neg\textsc{near}(x,y)]\}|$

5.2.1.2 The semantics of ****f:set****

I now incorporate GRAN(P) and COL(P) into the Subpredicative Iceberg Semantics model. Recall that in `FREE`* languages, like Mandarin Chinese, **f:set* always projects, and so the semantics of **f:set* is simply a closure under sum as in (17), repeated here as (28). The semantics of **f:set* in `PRESUP`* languages such as English, Welsh and Hejazi/Modern Standard Arabic on the other hand are a slightly different story, as

[17] Alternatively, one could take the route that Grimm (2012b) takes, and assume that universal quantification over a series of connected entities forms some offline judgement such that it is not available for online calculation of truth conditions.

**f:set* is associated with a set of presuppositions related to construed aggregation. With respect to The Order of Aggregation there are two potential realisations for construed aggregation presuppositions on **f:set*. The first is the English-type **f:set*, as in (29), where only granular aggregation is associated with **f:set*. Under these semantics, **f:set* will project when GRAN(P) is satisfied. For example, in the case of a filtered set built from a root such as CORN$^{\sqrt{c}}$, MILLET$^{\sqrt{c}}$, or BARLEY$^{\sqrt{c}}$, the GRAN(P) presupposition of **f:set* will be satisfied and a number neutral uncountable predicate will compose as a result. In the case of a filtered set built from a root like DUCK$^{\sqrt{c}}$, PEAR$^{\sqrt{c}}$, or BEE$^{\sqrt{c}}$, the GRAN(P) presupposition will not be satisfied, and **f:set* will not project, ultimately leading to a singular count predicate composition as the only possible derivation. The second version of **f:set* is the collective/singulative version found in Welsh and Hejazi/Modern Standard Arabic as in (30). Under these semantics, **f:set* will project when either GRAN(P) or COL(P) is satisfied. In this version of **f:set*, the derivation will compose when **f:set* is projected over *e.g.* DUCK$^{\sqrt{c}}$.

(28) `FREE`* languages (Mandarin Chinese):

$$\text{*}\textit{f:set} \rightarrow \lambda \mathrm{P}\left\{ \lambda x[{}^{*}P(x)] \right.$$

(29) `PRESUP`* languages, number marking language version (English):

$$\text{*}\textit{f:set} \rightarrow \lambda P\begin{cases} \lambda x[{}^{*}P(x)] & \text{if Gran(P) = TRUE} \\ \perp & \text{otherwise} \end{cases}$$

(30) `PRESUP`* languages, singulative language version (Welsh, Arabic):

$$\text{*}\textit{f:set} \rightarrow \lambda \mathrm{P}\begin{cases} \lambda x[{}^{*}P(x)] & \text{if Col(P)} \vee \text{Gran(P) = TRUE} \\ \perp & \text{otherwise} \end{cases}$$

With this, I now update the schematic in 5.10 to 5.11 which is overlaid with the presuppositions associated with The Order of Aggregation.

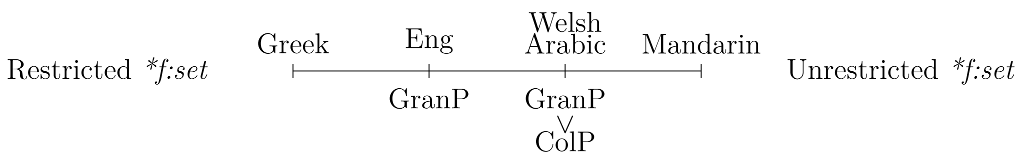


Figure 5.11: The scale of *f:set projection overlaid with The Order of Aggregation

At this point, note that English has a small set of number neutral uncountable terms which denote notional collective aggregates (fruits and vegetables), as in (31). Under the semantics in (29), the property *e.g.* FRUIT should not compose as a number neutral uncountable term as it does not satisfy GRAN(P).

(31) FRUIT, CELERY, SPINACH, ROCKET, KALE, ASPARAGUS, OKRA

I suggest that the scaled nature of *f:set is not restricted to the categories COL(P) and GRAN(P), and other presuppositions may well be relevant to *f:set in PRESUP* languages. However, due to the scaled nature of *f:set, if a PRESUP* language associates GRAN(P), but not COL(P) with *f:set, then any minor categories associated with *f:set that are not captured by GRAN(P) must be a subset of COL(P). In this manner, we may expect a small number of notional collective aggregate number neutral uncountable terms in English, as in (31), but we would not expect number neutral uncountable coding for non-notional collective aggregates, such as DOG.

Before closing this section, I will make clear a final assumption regarding the composition of i-sets. While I assume that the projection of *f:set in any PRESUP* language is based on implicit judgements, I assume that the subpredicative derivation of any i-set is automatic. Competent speakers do not freely switch between singular count and number neutral uncountable uses of a given term. It is not the case that *f:set in PRESUP* languages is a choice for any specific speaker, *i.e.* it is not the case that a Welsh speaker may stop projecting **f:set* for *e.g.* MOCH (pig.COL = *'pig(s)'*), essentially choosing with each use whether MOCH is a count or number neutral uncountable term. Rather, the judgements that underlie **f:set* projection should be taken as implicitly agreed upon by the community of speakers. In a way, once a language community has decided that **f:set* projects over a given i-set, it will always project over that given i-set, until the community of speakers decide otherwise.[18]

5.2.1.3 Section summary

In this section I have proposed the basic architecture for nominal terms under an enriched Subpredicative Iceberg Semantics model. The account follows the typical assumption that construed ontological judgements of speakers of different languages do not substantially differ. Humans are endowed with an innate, pre-linguistic, ability to distinguish between substances and objects. The Subpredicative Iceberg Semantics account suggests that in line with the mapping property all substances in all languages are stuff mass terms and have a single derivational path via the SUB filter. The manner in which a language may refer to objects (as either singular count or number neutral uncountable) is parametric, with the (lack of) projection of **f:set* being the locus of crosslinguistic variation. `FREE`* languages (*e.g.* generalised classifier languages like Mandarin Chinese) obligatorily project **f:set*, while `PRESUP`* languages endow **f:set* with a set of presuppositions which limit its projection. In accordance with The Order of Aggregation, `PRESUP`* languages may associate **f:set* with a granular aggregate presupposition (*e.g.* number marking languages like English) or with both a granular aggregate presupposition and collective aggregate presupposition (*e.g.* singulative languages like Welsh and Hejazi/Modern Standard Arabic).

[18] There is one exception to this: when languages have two terms for the same concept, one count, one number neutral uncountable. For example, English COW and CATTLE arguably share the same root concept and the difference can be understood as COW being a singular count term with no **f:set* projection, but CATTLE is a structure with **f:set*.

5.2.2 Accounting for exceptions

The semantics proposed for **f:set* allows for free composition of number neutral uncountable terms for any granular aggregate notion in English and any granular or collective aggregate notion in Welsh and Hejazi/Modern Standard Arabic. Yet, as discussed at length in §4.2.1, there are exceptions to any proposed mapping from concept to language for the number neutral uncountable class, both within and across languages. This includes construed aggregation. To recap, consider Table 5.1 and Table 5.2, which are repeated from §4.2.3 for convenience, which exemplify terms that are in the singular/plural class for Welsh and Hejazi/Modern Standard Arabic despite the fact that they refer to construed aggregate concepts. Why should this be? There must be an account for why in some cases **f:set* projection fails *despite* the aggregation presupposition being well defined. In this section, I explore the effects of animacy, size, and interaction on the composition of number neutral uncountable terms. The available data suggests that Welsh is sensitive to size-based considerations, while on the other hand Hejazi/Modern Standard Arabic is sensitive to animacy considerations.

	Arabic	**Gloss**
	niktaariin → niktaariin-aat	nectarine → nectarines
	butghuuth → baraghiith ▼	flea → fleas
	3nkabuut → 3nakib ▼	spider → spiders
Singular/	3qrab → 3qarib	scorpion → scorpions
Plural	dabbar → dabaabiir ▼	wasp → wasps
	ṣirṣar → ṣaraṣiir	cockroach → cockroaches
	khunfusa’ → khanaafis	beetle → beetles
	jundub → janadib	grasshoppers → grasshopper

▼ The equivalent term is collective/singulative in Welsh

Table 5.1: Exceptions to Arabic collective class

	Count term	Gloss
	almon → almon-au ▼	almond → almonds
	pecan → pecan-au ▼	pecan → pecans
	olif → olif-au ▼	olive → olives
	nectarîn → nectarin-au	nectarine → nectarines
	afocado → afocado-s ▼	avocado → avocados
	mango → mango-au	mango → mangoes
	ciwi → ciwïod ▼	kiwi → a kiwi
	afal → afal-au ★ ▼	apple → apples
	oren → orenn-au ▼	orange → oranges
	lemon → lemon-au ▼	lemon → lemons
	tomato → tomato-s ▼	tomato → tomatoes
	pupur → pupr-au ▼	peppers → a pepper
	banana → banana-s ▼	banana → bananas
Singular/ Plural	pomgranad → pomgranad-au ▼	pomegranates → a pomegranate
	grawnffrwyth → grawnffrwyth-au ★	grapefruit → grapefruits
	pîn-afal → pîn-afal-au ▼ ★	pineapple → pineapples
	ciwcymbr → ciwcymbr-au ▼ ▼	cucumber → cucumbers
	corbwmpen → corbwmpenn-i ★ ▼	courgette → courgettes
	melon → melon-au ▼	melons → a melon
	dyfrfelon → dyfrfelon-au ★▼	watermelons → a watermelon
	maro → maro-s	marrow → marrows
	pompiwn → pompiyn-au ▼	pumpkin → pumpkins
	miltroed → miltroed-iaid ★	millipede → millipedes
	sgorpion → sgorpion-au	scorpion → scorpions
	mosquito → mosquito-s ▼	mosquito → mosquitos
	locust → locust-iaid ▼	locust → locusts
	glöyn byw → glöynn-od byw ▼ ★	butterlfy → butterflies
	pili-pala → pili-palod ★ ▼ (S)	butterfly → butterflies
	rhesin-en → rhesin-s	raisin → raisins
	wnion-yn → nion-od ▼	onion → onions
	tat-en → tat-ws	potato → potatoes
	trog-en → trog-od ★	tick → ticks
Faux Singulatives	malw-en → malw-od (N) ★▼	snail → snails
	gwlith-en → gwlith-od ★▼	slug → a slug
	sionc-yn gwair → sionc-od gwair ★	grasshopper → grasshoppers
	chwil-en ddu → chwil-od duon ★	cockroach → cockroaches
	chwil-en → chwil-od ★	beetle → beetles

★ The term is *not* borrowed.

▼ The equivalent term is collective/singulative in Arabic

Table 5.2: Exceptions to the Welsh collective class

5.2.2.1 Animacy

Animacy is known to be linked to number marking. In Georgian, animate subjects experience plural verbal agreement, while inanimate subjects experience singular verbal agreement. Similar observations are recorded for Turkish nominal terms which take the -LER suffix, where inanimate plural subjects have singular verbal control. These observations led Smith-Stark (1974) to posit that languages differ on how they mark number with respect to how they 'split' THE SMITH-STARK ANIMACY HIERARCHY presented in Figure 5.12. The claim is that if a language splits the hierarchy at any point, number marking respects the split segments. For example, if a language splits the hierarchy between the *animate* and *human* categories, then the *inanimate* and *animate* categories are said to have one pattern of marking/agreement while *human, rational, kin, addressee* and *speaker* categories would have a different pattern of agreement marking.

inanimate < animate < human < rational < kin < addressee < speaker

Figure 5.12: The Smith-Stark Animacy Hierarchy

Smith-Stark posits that The Smith-Stark Animacy Hierarchy does not necessarily apply to all languages and language categories, but rather captures a strong preference for nominal number systems, and therefore counterexamples may arise. This claim was discussed by Corbett (1996), who argues that The Smith-Stark Animacy Hierarchy does not apply to the Maltese collective/singulative class due to the mixture of animate and inanimate terms within that class. More recently, Grimm (2012b, 2018) offers the reverse view for animacy and collective/singulative systems. For Grimm, a modified version of The Smith-Stark Animacy Hierarchy is taken to be the direct inverse the The Scale of Individuation (See §4.2.3.3). Similar to how Smith-Stark argued that languages have a plurality split at different points of The Smith-Stark Animacy Hierarchy, Grimm argues that collective/singulative languages differ on how high they ascend The Scale of Individuation/Animacy. Therefore, the higher an individual is in a measure of animacy, the more likely it is that the term that refers to that individual will be part of the collective/singulative class.

It turns out that animacy is a poor indicator of the collective/singulative class in Welsh. To see why, consider Table 5.3, which lists notional aggregates that are part of the collective/singulative class in both Welsh and Hejazi/Modern Standard Arabic. The first conclusion, under a Grimm style analysis is that lower animacy increases the probability that a term is in the collective/singulative class. Why, then, should the Arabic collective/singulative system be more productive than the Welsh system in the notional category of fruits and vegetables? Why should the collective aggregate terms in Table 5.2 be part of the Welsh singular/plural class, if low animacy is the key to the collective/singulative class? Why have these notional kinds escaped the collective/singulative class in Welsh, but not Arabic? It would seem that in Arabic there is a good coverage of inanimate concepts being part of the collective/singulative class, but in Welsh this coverage is disrupted. This is not to say that animacy does not play a role in the Welsh collective/singulative class. Rather, animacy alone is a poor indicator of the Welsh collective/singulative class specifically.

Welsh	**Arabic**	**Gloss**
mwyar → mwyar-en	tuut → tuut-ah	berry
grawnwin → grawnwin-en	3nb → 3nb-ah	grape
rhuddygl → rhuddygl-en	fujl → fujl-ah	raddish
moron → moron-en	jazar → jazr-ah	carrot
maip → meip-en	laft → laft-ah	turnip
cennin → cenhin-en	kurraat → kurraat-ah	leek
letys → letys-en	khas → khas-ah	lettuce
ceirios → ceirios-en	karaz → karaz-ah	cherry
dêts → det-en	tamr → tamr-ah	dates
bricyll → bricyll-en	moshosh → moshmosh-ah	apricot
eirin → eirin-en	barquuq → barquuq-ah	plum
ffigys → ffigys-en	tiin → tiin-ah	fig
gellyg → gellyg-en	’ijaaṣ → ijaaṣ-ah	pear

Table 5.3: Welsh and Arabic inanimate collective/singulative terms

5.2.2.2 Size

Animacy may not be strongly predictive of the collective/singulative class in Welsh, but size is. Consider the Welsh terms for the notional categories of *winged insects* and *wingless insects* in Table 5.4 and Table 5.5 respectively. Insects typically come in groups or swarms. Therefore, filtered sets associated with insect properties are able to fulfil the collective aggregate presupposition and compose as number neutral uncountable (collective) i-sets. Despite this, there are insect terms which have escaped the collective/singulative class.

I claim there is a size effect. The singular/plural terms in both the winged and wingless insects notional categories are terms which refer to individuals that are, on average, the largest in their respective categories. Conversely, the collective/singulative terms are the terms for individuals which are, on average, the smallest in their respective categories. Interestingly, the faux singulative pattern seems to apply to individuals that are middling in size, *i.e.* the faux singulative class refers to individuals which are typically larger than individuals denoted by collective/singulative terms, but smaller than individuals denoted by singular/plural terms. To this end, Table 5.4, and 5.5 are ordered for (average) size. It is not only the category of (winged) insects in Welsh that seems to have a size influence. Other categories with a size-influence include the categories of grains/seeds/pulses/nuts, berries, aggregate fruits, stone fruits, root vegetables and cruciferous vegetables. I refer the reader to the word lists in Appendix A to verify this claim.

	Welsh	**Gloss**
Collective/ Singulative	gwybed → gwybed-yn	gnats → a gnat
	clêr → cler-en	flies → a fly
	gwenyn → gwenyn-en	bees → a bee
	cacwn → cacyn-en	wasps → a wasp
Faux Singulatives	sioncyn gwair → sioncod gwair	grasshopper → grasshoppers
	chwilen ddu → chwilod duon	cockroach → cockroaches
	chwilen → chwilod	beetle → beetles
Singular/ Plural	mosgito → mosgito-s	mosquito → mosquitos
	locust → locust-iaid	locust → locusts
	glöyn byw → glöynn-od byw	butterfly → butterflies

Table 5.4: Welsh winged insect terms

	Welsh	**Gloss**
Collective/ Singulative	gwiddon → gwiddon-yn	mites → a mite
	chwain → chwann-en	fleas → a flea
	lleu → lleu-en	louse → a lice
	mogrug → morgrug-yn	ants → an ant
	myrion → myrion-en (S)	ants → an ant
	cynrhon → cynrhon-yn	maggots → a maggot
	lindys → lindys-yn	caterpillars → a caterpillar
	abwyd → abwyd-yn (S)	worms → a worm
	malwod → malwod-en (S)	snails → a snail
	pryfed cop → pryfed cop-yn (N)	spider → a spider
Faux Singulatives	trogen → trogod	tick → ticks
	malwen → malwod	snail → snails
	gwlithen → gwlithod	slug → a slug
Singular/ Plural	miltroed → miltroed-iaid	millipede → millipedes
	sgorpion → sgorpion-au	scorpion → scorpions

(S) = Southern Wales dialects, (N) = Northern Wales dialects

Table 5.5: Welsh wingless insect terms

Let us then assume that size is a relevant factor to Welsh number neutral uncountable composition in that entities construed as large may escape the collective/singulative class, even if those entities are defined for notional aggregation. If this is the case, it must come with the following caveats. First, as also argued by Wierzbicka (1988), any comparison of size between individuals must be relative to like-individuals. In the case of the examples at hand, (winged) insects must be compared against other (winged) insects. A second caveat is that if we consider a size-based explanation for certain notional concepts escaping the collective/singulative class then it does not follow that Welsh speakers necessarily construe *e.g.* caterpillars as small wingless insects while they construe *e.g.* millipedes as large wingless insects. Rather, when individuals which fulfil the construed aggregation presupposition are ordered for average size, there is an arbitrary size point at which a term entering the collective/singulative class fails. The final caveat is that any notion of size is not universal to all collective/singulative languages, as there is no evidence that Arabic is sensitive to size of entities with respect to the collective/singulative class.

Should size be kept as an influencing factor for the Welsh collective/singulative class if there are so many caveats? Some authors would think not. Grimm (2012b, 2018) for one is critical of size as a significant factor in the coding of the collective/singulative class. For Grimm (2012b), the collective/singulative class is defined via individuation, which, following experimental evidence inMiddleton et al. (2004), is taken as a conglomerate measure which treats construed aggregation as primary, but has influences of animacy and interaction. Grimm concludes that size observations are ‘epiphenomenal’ as less individuated entities tend to be small entities, and so size effects are captured by individuation. Similarly, Stolz (2001) is critical of a size-based analysis, aligning the collective/singulative class primarily with construed aggregation. Though both Grimm and Stolz are critical of a size-based analysis of collective/singulative systems, they are only critical of size as a *primary* source of collective/singulative systems. Stolz (2001) especially does not outright rule out that size may be important for the Welsh collective/singulative class stating:

> In these cases, size is of secondary importance. The perceptually shared and foregrounded feature between prototypical members once again is the characteristic way of living together in swarms, flocks, herds and shoals
> (Stolz, 2001, p.65, adapted)

In this way, I agree with Stolz (2001) and to an extent Grimm (2012b): size can only ever be a secondary factor in consideration for the (Welsh) collective/singulative class. However, I invert the argument: it is not construed *smallness* that makes a concept likely to be part of the collective/singulative class as a secondary judgement. Rather, construed *largeness* may prevent an otherwise perfectly defined aggregate concept from being represented as a number neutral uncountable term at the linguistic level. In terms of Subpredicative Iceberg Semantics, if the construed aggregation presupposition is satisfied for a filtered set, then **f:set* projects unless those individuals in the filtered set are construed as comparatively large.

Taking size considerations as a secondary condition for the Welsh collective/singulative class provides a straightforward account of Welsh dialectal variation. Though the inventory is small, consider the terms in Table 5.6, which show variation between Northern Wales and Southern Wales dialects on the representation of certain terms as part of the collective/singulative class or the faux singulative class. If the collective/singulative class is sensitive to an arbitrary size judgement then it stands to reason that dialects will differ where this arbitrary point is. For the examples in Table 5.6, it may be that snails, onions, and worms are sitting on the proverbial fence with their size being able to prevent their incorporation into the collective/singulative class. Interestingly, notice that the Northern Wales terms in Table 5.6 which are not in the collective/singulative class are nonetheless part of the *faux singulative* class. I have already suggested that faux singulative marking applies to those terms denoting individuals middling in relative size. This being the case, the fact that variation regards collective/singulative versus faux singulative coding, and not collective/singulative verses singular/plural coding, is not surprising. It would seem that when ordered for size, there is a zone of indifference where aggregate concepts may fall into either the collective/singulative or faux singulative class.

Collective → singulative	**Faux singulative → plural**	**Gloss**
winwns → winwns-yn (S)	nion-yn → nion-od (N)	Onion
abwyd → abwyd-yn (S)	mwyd-yn → mwyd-od (N)	Worm
malwod → malwod-yn (S)	malw-en → malw-od (N)	Snail

Table 5.6: Welsh collective/singulative dialectal variation

There is an interesting point to be made here regarding implicit grading. Sapir (1944, *et. seq.*) have shown that implicit grading is relevant to adjectives of size, with a `ZONE OF INDIFFERENCE`. The argument put forward here is that implicit grading of size is a mechanism that interacts with nominal structure, at least in Welsh, where construed large size may prevent a concept entering the collective/singulative class. Terms of middling size may then trigger faux singulative agreement for singular/plural terms. An alternative hypothesis which is compatible with size-gradable account is that faux singulative coding may also be a by-product of reanalysed singulatives. The idea is that faux singulatives historically been part of the collective/singulative class, but over time shifted to the singular/plural class while maintaining 'singulative' marking in the newly analysed singular form. An example where this has happened is the CWNING-EN/CWNING-OD (RABBIT/RABBITS) pair discussed in §2.1.5.

With this, I will take the explicit stance that relative size is a contributing factor to the collective/singulative class in Welsh. I do not take the influence of size to be epiphenomenal a la Grimm (2012b). Rather, I suggest that size is an implicit secondary consideration in the derivation of Welsh number neutral uncountable terms, where a measure of relative large size may prevent an aggregate concept from entering the collective/singulative class.

An immediate follow up question is why should this size have importance in Welsh, but not in Arabic? A tentative answer lies in the form of the singulative marker. In Welsh, the singulative marker -YN/-EN is homophonous not only with singular endings in suppletion patterns (faux singulatives) but also with one form of synthetic diminutive forms (Heinz, 2009; Nurmio, 2017, 2020; Rosiak, 2009, 2013; Stolz, 2001). I argue that this connection is responsible for a size based influence on **f:set*. Examples of -YN/-EN as a diminutive are in Table 5.7.

Singular	**Diminutive**	**Gloss**
sgwar	sgwar-yn	square → little square
tamaid	tameid-yn	piece → little piece
bachgen	bachgenn-yn	boy → little boy

Examples from Rosiak (2013)

Table 5.7: Welsh diminutive marking

The homophony of the Welsh diminutive and singulative is well versed, though there is no consensus on interactions between the singulative and diminutive functions, if such interactions exist at all. Stolz (2001, p.64) suggests that while diminutives and singulatives are semantically separate, there may be some diachronic or conceptual link between the two. In a similar manner, Nurmio (2020) takes singulative and diminutive markers to be synchronically distinct. On the one hand, Heinz (2009) proposes that -YN/-EN is a multifunctional morpheme which acts as both a diminutive and a singulative dependent upon the base it attaches to. Similarly, Rosiak (2009, 2013) takes the diminutive function of -YN/-EN as primary and the singulative function as secondary.

The history and the connection between diminutives and singulatives is not the focus of this work, but it does stand to reason that the Welsh singulative marker is in some way connected to diminutivisation, even if the connection is a reanalysis of coincidental development. It is not far fetched, I surmise, to propose that if a singulative is connected to diminutivisation (even if via reanalysis), then the entire collective/singulative system which it is part of may be linked to smallness. This is not to say that singulatives are diminutives. Rather, I suggest that the Welsh singulative class has adopted a notion of smallness adjacent to the the primary function of division, and this notion of smallness is reflected in the ability for construed relative largeness as able to prevent **f:set* projection.

It remains to be seen if this suggestion is crosslinguistically viable or Welsh specific. It is known that Welsh is not the only language said to have a connection between singulative and diminutive markers. Kagan and Nurmio (2024) claim there is a systematic typological overlap between singulative markers and diminutive markers.[19] Only thorough testing of the collective/singulative inventories of other singulative/diminutive linked languages will reveal if my

[19] In fact, Mathieu (2012a) connects both the Ojibwe diminutive and singulative under the same morpheme. Yet, despite there being overlap in functions, Kagan and Nurmio (2024) are careful in explaining that singulativisation does not necessarily entail diminutivisation: Kagan and Nurmio (2024); Kagan (2024) give an analysis of Russian suffix -INK as a composition of a diminutive and singulative morphemes.

suggestion spans beyond Welsh. However if the proposal that a link between diminutive and singulative morphemes results in secondary size construals for the collective/singulative class is on the right track, then we would, correctly, not expect the Arabic collective/singulative system to be secondary linked to smallness at all, as the Arabic singulative marker is not systematically connected to diminutivisation.[20]

5.2.2.3 A variable measure of individuation

In the extended Subpredicative Iceberg Semantics approach being built here, so far construals of aggregation are both necessary and sufficient for number neutral uncountable derivation in `PRESUP*` languages. Through the subsequent discussion, it emerges that number neutral uncountable composition is also sensitive to factors of animacy and size, to different measures, in Welsh and Arabic.

Ultimately, both measures of size and animacy are sub-measures of individuation in the Grimm (2012b); Middleton et al. (2004) and Wierzbicka (1988) sense. I propose that while construals of aggregation are necessary and sufficient for number neutral uncountable composition, there is a secondary `MEASURE OF INDIVIDUATION` (`MOI`) associated with **f:set* projection which further restricts legal projection. The `MOI` is modelled as in (32), where MOI(P) is the individuation measure of a filtered set and $\boldsymbol{n}$ is an extralinguistic standard for individuation for which this measure is compared. MOI(P) should be understood as a conglomerate measure of features relevant to perceptual individuation with respect to the filtered set. At a minimum,

[20]Though *c.f.* Jaradat and Jarrah (2022, p.516) who in a footnote states that in Syrian and Egyptian dialects the diminutive -AYAH may be used for individuation (*i.e.* function as a singulative), but they restrict its use grains, fruits and vegetables while the more standard version of singulativisation is the standard singular feminine marker -AH. In this way, singulativisation *is* connected to diminutivisation, at least in this one function. Note however, -AYAH is perhaps a composition of [singulative+diminutive]. Though not not the aim of the work, Alshboul et al. (2013) happen to have examples of diminutives that apply to singulative forms clearly showing -AYAH as a [singulative+diminutive] composition *e.g.* SHAJAR-AH (tree.COL-SING = *'tree'*) → SHUJ-AY-AH (tree.COL-SING.DIM = *'little tree'*). I leave the discussion of whether Arabic singulatives are indeed related to diminutivisation, but insofar as I can see, the process mirrors what Kagan (2024); Kagan and Nurmio (2024) argue for Russian: it is compositional, and singulativisation is not inherently linked to diminutivisation.

MOI(P) calculates a value based on submeasures of `SIZE` and `ANIMACY` of the entities in the filtered set. If it is the case that the MOI(P) is larger than the extralinguistic standard, **f:set* does not project, ultimately leading to singular count i-set derivation. With this, the semantics for **f:set* are updated in (33) for singulative languages and (34) for number marking languages.

(32) `MEASURE OF INDIVIDUATION` (`MOI`)
MOI(P) < $\boldsymbol{n}$
Where P is a filtered set

(33) `PRESUP`* languages, singulative language version (Welsh, Arabic):

$$\textit{*f:set} \rightarrow \lambda \mathrm{P} \begin{cases} \lambda x[{}^{*}P(x)] & \text{if ((Col(P)} \vee \text{Gran(P) = TRUE) and MOI(P) < n)} \\ \perp & \text{otherwise} \end{cases}$$

(34) `PRESUP`* languages, number marking language version (English):

$$\textit{*f:set} \rightarrow \lambda P \begin{cases} \lambda x[{}^{*}P(x)] & \text{if ((Gran(P) = TRUE) and MOI(P) < n)} \\ \perp & \text{otherwise} \end{cases}$$

Languages may differ on the specifics of MOI(P) calculation: both animacy and size may contribute to the measure of MOI(P) but they need hold the same weight of importance across all languages. This idea is schematised in (35). Welsh and Hejazi/Modern Standard Arabic are cases in point. For Welsh, a measure of size is weighted more than than a measure of animacy in the calculation of MOI(P), and so size effects are seen in the collective/singulative system where the (comparatively) smallest aggregate concepts enter the collective/singulative class. In Hejazi/Modern Standard Arabic, there are minimal size effects, if any. The measure of animacy is weighted more than the measure of size in MOI(P) calculation, and so (comparatively) large aggregate concepts may still enter the collective/singulative class. This is not to say that animacy considerations may not be taken into account in the calculation of

MOI(P) in Welsh, nor does it take size to be irrelevant in the calculation of MOI(P) in Hejazi/Modern Standard Arabic. Rather, these considerations are ranked differently for MOI(P) calculation.

(35) Languages may calculate MOI(P) in different manners:
Welsh MOI(P): Animacy < Size
Arabic MOI(P): Size < Animacy

If languages may vary on the calculation of MOI(P), it is reasonable to ask whether there is a canonical MOI(P) calculation. I have claimed that the Welsh collective/singulative class is linked to smallness only due to homophony of singulative and diminutive morphemes. I do not claim that the Arabic collective/singulative class is linked to animacy due to a similar homophonous relation. I tentatively suggest that the canonical MOI(P) calculation is the Hejazi/Modern Standard Arabic order where an animacy measure outweighs a size measure. The main reasoning for this is that animacy has proven, via The Smith-Stark Animacy Hierarchy, to be a fairly solid predictor of number marking patterns in other areas of number marking, and so MOI(P) is taken to be no different in this regard.

(36) `CLAIM:`
The canonical calculation of MOI(P) is Size < Animacy
However, languages where the singulative is linked to size (*e.g.* via homophonous singulative) are likely to prioritise size in their calculation of MOI(P).

The adoption of MOI(P) explains idiosyncratic minimal pairs that are difficult to explain under an animacy alone approach. For example, PLANT (child.COL = *'child(ren)'*) is collective/singulative in Welsh, but TIFL (child.SG = *'child'*) singular/plural in the dialects of Arabic. Children are (usually) small humans, and so the appearance of PLANT in the Welsh collective/singulative class falls out easily from a size-based approach.[21] On the other hand, children are higher animates, and so the fact that TIFL is not in the Hejazi/Modern Standard Arabic collective/singulative class is expected.

Before closing this section, there is one point to make abundantly clear: I take MOI(P) to be purely linguistic. The ordering of measures of individuation does not necessarily reflect on *perceptual* notions of individuation. In theory, two languages with disparate communities of speakers may conceptualise individuation of specific entities in exactly the same way, but their languages may differ due to how MOI(P) is calculated, and therefore the languages will differ on whether a term is in the collective/singulative class or the singular/plural class.

The suggestion that the MOI(P) is language specific turns out to be compatible with Middleton et al. (2004)'s experimental results. Recall the discussion in §4.2.3.1. Middleton et al.'s experiment results suggested that size is relevant for perceptual distinguishability but not the number neutral uncountable nominal class. This is the view taken here: size may or may not be linguistically relevant. An interesting consequence of this view is whether Middleton et al.'s results hold across languages. Recall that Middleton et al.'s experiment tested participants' coding of count and mass construals in an experimental setting in *English*. Considering MOI(P) calculation as intrinsic to number neutral uncountable composition, it very much could be the case that Middleton et al.'s experimental results arise due to the unconscious consideration of multiple construals de-emphasising size comparisons in the English calculations of `MOI`. A fruitful exploration going forward would be to see if different results abound for Welsh, *i.e.* whether the Welsh size effect proposed here can be induced in experimental conditions.

[21] The smallness of children playing a role in the collective/singulative status of PLANT is not a new argument. Stolz (2001) argues that smallness plays a role in the coding of PLANT as collective/singulative. Specifically, Stolz speculates that children being part of a larger group (members of a family) *and* smallness/endearment play a role in the collective/singulative nature of PLANT. However, *c.f.* Nurmio (2017, 62) who speculates that PLANT is a collective/singulative term due to its etymological roots. Nurmio points out that Latin PLANTA has horticultural meaning, linking PLANT to other collective/singulative terms with similar semantic properties.

5.2.2.4 Interaction

A reasonable question that follows is whether there are other sub measures of MOI(P). The first candidate for a measure that contributes to MOI(P) is `INTERACTION`. The effects of interaction and number neutral uncountable terms were discussed at length in §4.2.3. To recap, Wierzbicka (1988) was early to propose that interaction is relevant to individuation, a claim later tested experimentally and confirmed by Middleton et al. (2004), and adopted formally by Grimm (2012a,b). The basic conclusion is that even if a term denotes individuals which are defined for construed aggregation, the term may still appear in the singular/plural class due to interaction typically being with the individual rather than the group.

Interaction seems to play a role in nominal composition for terms which denote social, herd, pack, and farmed animals. Terms in these notional classes arguably denote aggregate individuals, as these animals either by nurture or nature typically come in herds, flocks, packs, prides, schools *etc.*. If we follow an interaction style approach (a la Wierzbicka, 1988; Middleton et al., 2004; Grimm, 2012b), we may be able to suggest that while construed aggregation in some sense is fulfilled, many of these terms are singular/plural due to the individuals being interacted with individually. Indeed, in a footnote Grimm (2018) speculates:

> Since there is more interaction with individual cows (*e.g.* milking) than with pigs or rabbits, cows would be more likely to be treated as individuals and be categorised in the singular/plural class.
> Grimm (2012b, 86)

I suggest an opposite view to Grimm.[22] Consider the farmed, social, and pack animal terms in Tables 5.8 and 5.9 for Welsh and Arabic, respectively. In some

[22] An interesting view compatible with Grimm's view comes from Tiersma (1982), who states that hens and cows may be come in groups, but roosters and bulls do not, as it is common practice to keep a single rooster or bull within the flock/heard. Similar reasoning is used by Tiersma (1982, p.835) in a footnote to explain why the aggregate concept *goose* seemingly has higher frequency counts in the singular than the plural (Tiersma generally predicts that terms for aggregate concepts are more frequent in plural use). For Tiersma, the difference comes down to city-living leading to fewer interactions with the group when compared to traditional rural living.

sense, each animal denoted comes in groups, and so in theory their related filtered sets fulfil the construed aggregation presupposition of **f:set*. Note also that many of these animals are higher animates, many of them large in size. The animacy and size factors are taken into consideration of the calculation of MOI(P) for each term, and may well provide a measure large enough so **f:set* does not project. However, consider the following: if a measure of interaction is also part of the calculation of `MOI`, then in some cases interaction with the *group* decreases individuation. In other words, an aggregate concept may be animate and large, but if there is considerable group interaction, then individuation may still be low, and **f:set* projects.

The consequence of this view is that due to the `MOI` the default state of affairs is that large/higher animate aggregate concepts (*e.g.* social/farmed animals) default to the singular/plural class, and their appearance in the collective/singulative class is the exception case. A desirable consequence comes from this: we may only expect aggregate (large) higher animates that are actually *interacted* with on the group rather than individual level to enter the collective/singulative class. This seems to be the case: as far as I am aware, the higher animates that do enter the collective/singulative class are farmed and/or domesticated.[23]

[23]Recall also Middleton et al. (2004)'s experiment where interaction with individuals of an aggregate increased the likelihood of assigning count status to its novel term. This seems to be the opposite effect of the farm-animal scenario: the `MOI` is raised enough to prevent **f:set*.

	Welsh	**Gloss**
Collective/ Singulative	pysgod → pysgod-yn	fishes → a fish
	adar → ader-yn	birds → a bird
	hwaid → hwyad-en	duck → ducks
	llygod → llygod-en	mice → a mouse
	moch → moch-yn	pig → pigs
Singular/ Plural	cyw→ cyw-ion	chicken → chickens
	gŵdd → gwydd-au	goose → geese
	alarch → elyrch	swan → swans
	colomen → colomenn-od	pigeon → pigeons
	dafad → defaid	sheep → sheeps
	gafr → geifr	goat → goats
	buwch → buch-od	cow → cows
	ceffyl→ ceffyl-au	horse → horses
	blaidd → bleiddiaid	wolf → wolves
	llew → llew-od	lion → lions
	eliffant → eliffant-od	elephant → elephants

Table 5.8: Welsh social, herd, pack, and farmed animal terms

	Arabic	**Gloss**
Collective/ Singulative	samak → samak-ah	fishes → a fish
	ṭa'iir → ṭa'iir-ah	birds → a bird
	baṭ → baṭ-ah	duck → ducks
	dajaaj → dajaaj-ah	chicken → chickens
	buum → buum-ah	swan → swans
	hamaam → hamaam-ah	pigeon → pigeons
	baqar → baqar-ah	cow → cows
Singular/ Plural	khinziir → khanaaziir	pig → pigs
	ghanam → aghnaam	sheep → sheeps
	maa3iz → mawaa3iz	goat → goats
	ḥiṣaan → ḥuṣun	horse → horses
	dhi'b → dhi'aab	wolf → wolves
	asad → usuud	lion → lions
	fiil → fiyalah	elephant → elephant

Table 5.9: Arabic social, herd, pack, and farmed animal terms

A final candidate for a sub-measure of MOI(P) is a tentative sub-notion of interaction which I term `MOVABILITY OF INANIMATES`, which is intended as a measure of how easy it is to physically pick up and move an individual. For example, apples are more movable than trees, as it is easy to pick up and move an apple compared to a tree. Trees are not movable, at least not without extended effort that involves prolonged interaction with axes and chainsaws. Moveability is seemingly important in the calculation of MOI(P), as the notional category of *trees* in both Welsh and Arabic has (to my knowledge) 100% collective/singulative coverage, a feat not seen in any other notional category. I am not aware of any tree term in Welsh that is a singular/plural term. In this way, tree terms fulfil the collective aggregate presupposition of **f:set*, and their MOI(P) calculation is such that it does not constitute a measure large enough to prevent **f:set* projection.

Welsh	**Gloss**
coed → coed-en	trees → a tree
gwern→ gwern-en	alders → an alder
onn → onn-en	ashes → an ash
bedw→ bedw-en	birches → a birch
ysgaw → ysgaw-en	elders → an elder
cyll → coll-en	hazels → a hazel
celyn → celynn-en	holly tree → a holly tree
derw → derw-en	oaks → an oak
cerddin→ cerddin-en	rowans → a rowan
helyg→ helyg-en	willows → a willow
sycamorwydd→ sycamorwydd-en	sycamores → a sycamore
castanwydd → castanwydd-en	chestnut trees → a chestnut tree
pinwydd → pinwydd-en	pine trees → a pine tree
ffynidwydd → ffynidwydd-en	fir trees → a fir tree
pefrwydd → pefrwydd-en	spruce trees → a spruce tree

Table 5.10: Welsh tree terms

Arabic	**Gloss**
shajar → shajar-ah	trees → a tree
bunduq → bunduq-ah	hazel trees → a hazel tree
ṣafṣaaf→ ṣafṣaaf-ah	willows → a willow
zaytuun → zaytuun-ah	olive trees → olive tree
sadr → sadr-ah	lotus trees → lotus tree
ṭalḥ → ṭalḥ-ah	banana trees → a banana tree
ghaar →'ghaar-ah	laurel trees → laurel tree
mishmish → mishmish-ah	apricot trees → apricot tree
nakhl → nakhl-ah	palm trees → palm tree

Table 5.11: Arabic tree terms

5.2.2.5 Borrowing

I have so far suggested that the semantics of number neutral uncountable composition necessitates that the individuals denoted by a term are typically construed as aggregate individuals, but other factors such as size, animacy, and manner of interaction influence number neutral uncountable composition. However, as many collective/singulative exception terms in Welsh are borrowed from English, there is perhaps a temptation to argue that the exceptions in the Welsh collective/singulative class is the result of a borrowing effect.

I follow Stolz (2001) in taking the borrowing effect to be minor, as there are multiple borrowings that are steadfastly in the collective/singulative class, as shown in Table 5.12.[24] This being said, minor borrowing effects may explain some exceptions. For example, BLODFRESCH is a collective/singulative term which may refer to either broccolis or cauliflowers. The borrowed terms BROCOLI (broccoli.SG = *'broccoli'*) and COLIFFLŴAR (cauliflower.SG = *'cauliflower'*) on the other hand are singular/plural terms. It may very well be the case that COLIFFLŴAR and BROCOLI are blocked from entering the collective/singulative class (*i.e.* projection of **f:set*) due to their being no lexical gap to be filled.

[24]Incidently, the borrowed terms in the Welsh collective/singulative class denote comparatively small individuals, indicating that size takes precedence over borrowing, a welcome result.

	Welsh	**Gloss**
Collective/ Singulative	pys → pys-en	peas → a pea
	grêps → greps-en	grapes → a grape
	winwns → winwnsyn (S)	onions → an onion
	letys → letys-en	lettuces → a lettuce
	dêts → det-en	dates → a date
	plwmws → plwmwns-en (S)	plums → a plum
	ffigys → ffigys-en	figs → a fig
	pêrs → per-en (N)	pears → a pear
Faux Singualtives	rhesin-en → rhesin-s	raisin → raisins
	tat-en → tatw-s	potato → potatoes
Singular/ Plural	almon → almon-au	almond → almonds
	pecan → pecan-au	pecan → pecans
	coliffl ŵar → coliffl ŵars	cauliflower → cauliflowers
	olif → olif-au	olive → olives
	nectarîn → nectarin-au	nectarine → nectarines
	mango → mango-au	mango → mangoes
	lemon → lemon-au	lemon → lemons
	tomato → tomato-s	tomato → tomatoes
	sgorpion → sgorpion-au	scorpion → scorpions
	mosgito → mosgito-s	mosquito → mosquitos
	locust → locust-iaid	locust → locusts

Table 5.12: Welsh borrowed terms across countability classes

5.2.2.6 Section summary

In this section I have proposed that number neutral uncountable composition in PRESUP* languages involves not only construed aggregation presuppositions (for natural kind denoting entities), but also a judgement of individuation, MOI(P), which ensures number neutral uncountable composition abounds in the cases where individuals denoted by the filtered set have a low measure of individuation. The calculation of MOI(P) is taken to be a conglomerate measure of submeasures of size, animacy, and interaction factors. While MOI(P) is based on conceptual factors, it is inherently linguistic: languages may differ on the exact recipe of the calculation of MOI(P), with Welsh (but not Arabic) placing greater importance on submeasures of size over animacy.

5.3 Discussion

The account put forward in this chapter has used the Subpredicative Iceberg Semantics framework to connect the wisdom from lexico-meaning accounts (that the count/mass distinction is not arbitrary) with the crosslinguistic simplicity afforded to constructionist accounts. Expanding on the minimal compositional elements of de Vries and Tsoulas (2021)'s Subpredicative Iceberg Semantics, there are four subpredicative nominal i-set derivations, as in (37). This chapter has focussed on singular count, stuff mass and number neutral uncountable i-set derivation. Note, I discuss number neutral countable predicates in more detail in chapter 6.

(37) NOMINAL TYPES IN SUBPREDICATIVE ICEBERG SEMANTICS

Singular count	⟨**body**, **base**⟩	
Number Neutral countable	⟨***body**, **base**⟩	
Number neutral uncountable	⟨***body**, ***base**⟩	
Stuff mass	⟨**body**, **base**⟩	where the **base** is not disjoint

With this minimal architecture of nominal number, there are a number of testable predictions. The first prediction concerns mapping from construed ontologies to linguistic ontologies, and is related to Chierchia (2010b)'s mapping property. That is, substance mass terms have only one available i-set representation in and across all languages, leading to Generalisation One.

(38) GENERALISATION ONE: THE MAPPING PROPERTY

All substances, in all languages, map substance concepts to ⟨**body**, **base**⟩ i-sets where the **base** is disjoint (stuff mass).

Though all substance denoting properties ultimately compose as stuff mass terms, it is not the case that all properties denoting pre-theoretical individuals must be singular count. Indeed, $\langle$***body**, ***base**$\rangle$ terms denote pre-theoretical individuals.

The extended Subpredicative Iceberg Semantics approach proposed here for number neutral uncountable terms is close to Grimm (2012a,b, 2018)'s lexico-meaning account in spirit. Like Grimm, I suggest that construed ontological judgements of aggregation, individuation and interaction play a role in the composition of number neutral uncountable terms. The major difference between this account and Grimm (2012b)'s account is the implementation of how these construals interact with the grammar. For one, Grimm (2012b)'s account is lexico-meaning in nature while Subpredicative Iceberg Semantics is constructionist in nature. This allows for Subpredicative Iceberg Semantics to capture the similarities between object mass, collective and general number terms in a way that is not immediately clear in Grimm (2012b)'s account. Explicitly, object mass, collective and general number terms are all taken to be fundamentally the same type in Subpredicative Iceberg Semantics, but in Grimm's account, there is no discussion of how generalised classifier languages are incorporated into the system.

Secondly, in Grimm's account emphasis is placed on both animacy and cultural variation as predictive of the collective/singulative class. However, the Welsh collective/singulative class is not clearly predicted by animacy. Therefore, under Grimm's view, we must argue that cultural variation plays a significant role in ways that are increasingly unfalsifiable, as we must posit that *e.g.* Hejazi/Modern Standard Arabic speakers and Welsh speakers differ substantially on their construals for terms which belong to differing nominal number classes.

The extended Subpredicative Iceberg Semantics account proposed in this chapter offers an alternative solution. I suggest that `PRESUP`* languages necessarily associate **f:set* with a construed aggregation presupposition in the notional natural kind domain. The construed aggregation presupposition at a minimum is associated with granular aggregation (number marking languages), but may also optionally be associated with collective aggregation (singulative languages). Secondly, there is a compulsory conglomerate measure of individuation, MOI(P), which must have a low value relevant to some contextually salient measure else **f:set* will not project. While MOI(P) is based on construals of the world it is an intrinsically linguistic device. Languages are free to vary the credence afforded to sub-measures of MOI(P), which include at a minimum size, animacy, and interaction. In this way, we may say that Welsh speaking and Hejazi/Modern Standard Arabic speaking communities do not necessarily construe entities differently, but rather the languages prioritise different submeasures of individuation in the calculation of MOI(P), which in turn influences the projection of **f:set*. This delicate interplay manifests in Welsh, but not Arabic, having a secondary size consideration for the collective/singulative coding, while at the same time not positing that the speakers of these languages differ on their construals of the world. This is the main big-picture difference between Grimm's account and the one proposed here: the mechanics in Subpredicative Iceberg Semantics allow for crosslinguistic variation without heavy reliance on cultural variation.

(39) `GENERALISATION TWO: CROSSLINGUISTIC APPEARANCE OF NUMBER NEUTRAL UNCOUNTABLE TERMS`

Collective, object mass and general number terms are number neutral uncountable terms, *i.e.* ⟨***body**, ***base**⟩ compositions composed via **f:set* projection in the subpredicative structure.

a. Languages differ on their propensity to project **f:set*.
   (i) `FREE`* languages obligatorily project **f:set*
   (ii) `PRESUP`* languages associate **f:set* with a set of construed aggregation presuppositions and a MOI(P) measure which restrict its projection.

The crux of this account is an assumption that while there may be cross-cultural differences, humans do not fundamentally differ on their construals of the world. The languages we speak, however, differ on which construals are relevant for different structures in a non-arbitrary scaled way. As it turns out, there is experimental evidence from Gathercole et al. (2000) which supports this scaled.

In their experiments, Gathercole et al. aimed to determine underlying mechanisms in children's extension of novel words to collections. The experiment focused on Spanish monolingual speakers, English monolingual speakers and Welsh monolingual speakers.[25] These languages were chosen because Spanish is said to not have a count/mass distinction, while English has a rich object mass system, and Welsh has an even richer collective/singulative distinction. Gathercole et al. were interested in whether children relied on innate biases, specifically the whole-object bias (Markman, 1990), in the learning of new words, or whether competency of grammatical structure guides the learning of new words.

In Gathercole et al.'s first experiment, adults and children were shown nine sets of items drawn on paper. Each item was introduced to participants using a novel term in mass/count/singular/plural neutral syntax so as not to bias judgements, *e.g.* in the English child-condition: "ON THIS PAPER YOU CAN SEE MY BLICKET" (Gathercole et al., 2000, p.65). An example of the initial stimulus in the `ONE-MANY` condition is shown in Figure 5.13, left. Following this, participants were shown the choice conditions, exemplified in Figure 5.13, right, and were asked to choose which of the choices the novel word (BLICKET) refereed to, *e.g.* "WHICH OF THESE IS THE BEAR'S BLICKET?" (Gathercole et al., 2000, p.67). The `ONE-MANY` condition exemplified in Figure 5.13 specifically examined whether speakers extended the novel term to individual items or to several. The results showed that Spanish speaking participants picked the single item significantly more often than the Welsh speaking participants, and the English speaking participants performance was between the two groups.

[25]Gathercole et al. (2000) are careful to note that Welsh is typically spoken by bilingual communities, and considered participants monolingual Welsh if it was asserted that they spoke Welsh more than 80% of the time.

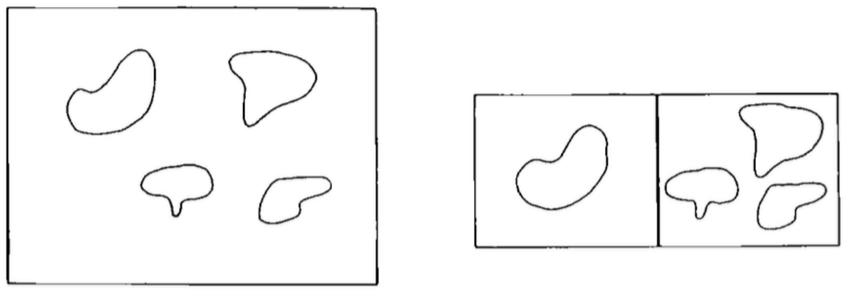


Figure 5.13: Gathercole et al.'s experimental pairs

In a second experiment, Gathercole et al. (2000) improved the stimuli such that they appeared in context of a bedroom alongside familiar items (a ball, a bed *etc.*). In this second experiment, Gathercole et al. found that children extended novel words in ways consistent with adult usage. By age two, Spanish speaking children showed a preference for novel terms to refer to 'one', English children by age four showed a preference for novel terms to refer to 'one', whereas Welsh speaking children by age four preferred to extend the novel noun to a 'many' interpretation, *i.e.* to collections.

Gathercole et al. take the view that there is little evidence for cognitive differences between speakers of languages and how they perceive the world (see §4.1.2), and yet children who speak languages with differing collection representation show different countability biases in word extension tasks. Gathercole et al.'s evidence suggests that while children and adults may not differ on their perceptual categorisation, they use tacit knowledge of how their languages categorise nominal expressions to categorise novel words. With this, Gathercole et al. suggest:

> The results for Spanish-speaking vs. English-speaking participants suggest perhaps a stronger pull among Spanish speakers in the direction of a whole-object approach than among English speakers.
> Gathercole et al. (2000, p.83)

These results are clearly compatible with the Subpredicative Iceberg Semantics model built in this chapter, where I explicitly take differences in number classes to be linguistic in nature, even though the underlying mechanisms are conceptual. Under the Subpredicative Iceberg Semantics account, and assuming The Lexical Statistics Hypothesis (Gleitman and Papafragou, 2005; Samuelson and Smith, 1999), then Gathercole et al.'s results have a simple theoretical explanation: speakers have tacit knowledge of the structures of their language and use this to categorise new words. The tacit knowledge in question is exactly the presuppositions associated with **f:set* and the calculation of the submeasures that constitute MOI(P).

Before closing this chapter, I call attention to one final point. The Order of Aggregation and the presuppositions associated with **f:set* discussed cover only notional natural kinds. I have offered no discussion of artificial collections (*e.g.* FURNITURE, KITCHENWARE), abstract concepts (*e.g.* DREAMS, FEAR), nouns with referents who do not actually exist in the real world (*e.g.* SANTA CLAUS), or exist at the construal level (*e.g.*. THE FIRST BABY TO BE BORN IN 2030. While such nouns are not the focus of this thesis, I suggest the following observations. First, the the IND filter is defined only for the physical domain. To account for abstract concepts, either a different filter must be proposed, or the formalisation of the IND filter must be expanded. Either way, the interaction of this filtered set and **f:set* must somehow be accounted for. Whether there is a conceptual reason for why DREAMS and FEAR fall into different countability classes is yet to be seen. Second, to account for artefact object mass terms in English such as FURNITURE, it very much could be the case that the the IND filter is used, but adjacent to The Order of Aggregation there is also an artificial collective aggregate presupposition associated with English **f:set*, ensuring number neutral uncountable composition. If this is the case, then artificial collective aggregate presupposition association with **f:set* is parametric: French, for example, would not associate **f:set* with a collective aggregate presupposition. Third, I assume reference to specific individuals applies at a compositional level distinct to subpredicative composition (see *e.g.* Landman (2020)'s semantics for definiteness).

The beauty of the Subpredicative Iceberg Semantics approach proposed here is that capturing crosslinguistic variation then becomes a matter of pattern searching. This is because the mechanism for the basic underlying primitives that derive number neutral uncountable terms across languages is minimal (the projection of **f:set*). In other words, the projection of **f:set* is the locus of linguistic variation, and while the specifics for its projection vary crosslinguistically, its projection nonetheless captures the fundamental grammatical primitives that underlay number marking languages, singulative languages, and generalised classifier languages. From here, the crosslinguistic differences are easy to incorporate: it is simply a matter of identifying the triggers for **f:set* in `PRESUP`* languages.

5.4 Summary and next steps

Theories of nominal number broadly fall into two camps, which are complementary in their strengths: the lexico-meaning and the constructionist approach. Lexico-meaning approaches typically shine in their ability to map construed ontologies to linguistic ontologies such that construed perceptual fact directly influences nominal number. While broadly successful, lexicalist approaches suffer in their ability to predict intra and cross linguistic variation. Constructionist accounts on the other hand are better equipped to account for general crosslinguistic variation, but leave little room for perceptual factors. The extended Subpredicative Iceberg Semantics approach presented here combines the strengths of lexico-meaning and constructionist approaches to nominal number such that there is a direct connection between construed ontological judgements and linguistic ontologies, while the account enjoys crosslinguistic predictive power.

In Subpredicative Iceberg Semantics, inspiration is taken from the constructionist tradition that the locus of variation for number categories is minimal and parametric. Specifically, the extension to Subpredicative Iceberg Semantics developed here maintains de Vries and Tsoulas's position that (the lack of) projection of **f:set* is the locus of crosslinguistic variation regarding count versus number neutral uncountable terms. The novel contribution proposed here is the exact nature of **f:set* in `PRESUP`* languages. In `FREE`* languages, like Mandarin Chinese, **f:set* is free and always projects. In `PRESUP`* languages, like English, Welsh and Arabic, **f:set* is associated with a set of construed aggregation presuppositions. In number marking languages, such as English, the only construed aggregation presupposition associated with **f:set* is that individuals in the related filtered set are granular aggregates. In singulative languages, such as Welsh and Hejazi/Modern Standard Arabic, the construed aggregation presuppositions associated with **f:set* are that individuals in the related filtered set are granular or collective aggregates. In all `PRESUP*` languages, **f:set* is associated also with a measure of individuation, MOI(P), such that a large measure of individuation results in **f:set* failing to project. With this, crosslinguistic variation of the terms which fall in the count or number neutral uncountable class is accounted for.

Chapter 6

Building a system: Classification and Plurality

This chapter offers an account of functions of number, including a semantics for individuating singulatives (as uncountable classifiers), packaging singulatives, countable classifiers, countable plurals, and uncountable plurals. The semantics are modelled in such a way that they are compatible with a split DP hypothesis of nominal number, *i.e.* where number functions are distributed across the nominal spine. The semantics proposed in this chapter accounts for why the plural of the singulative (including paucal of the singulative) is available in Hejazi/Modern Standard Arabic, but is completely absent in Welsh. For Hejazi/Modern Standard Arabic, the claim is that the Arabic sound plural of the singulative is a projection of paucal number while the broken plural of the singulative is the 'true' plural of the singulative. For Welsh, the semantics are such that the plural of the singulative is ruled out on Economy/vacuous grounds. The chapter concludes with light comments on deriving the interpretations of inclusive and exclusive plurality in a way compatible with (scalar) implicature approaches to plural interpretation.

In §6.1, I recap the basics of the semantics of i-set composition and set my assumptions regarding plural marking with respect to its interpretation in numeral and bare contexts.

In §6.2, I propose the semantics for basic functions of nominal number. Two types of plurals are discussed: countable plurals and uncountable plurals. English and Arabic are proposed to have countable plurals only, while Welsh is proposed to have an overt non-countable plural, but in numeral contexts a covert countable plural is induced as last resort. The semantics proposed for the Welsh overt plural in essence derives an i-set with the same underlying countability properties as a collective (number neutral uncountable) term. Alongside the plural semantics proposals, the semantics for three types of individuation are proposed: countable classifiers, non-countable classifiers, and portioning devices. The individuating singulative in Welsh and Hejazi/Modern Standard Arabic is taken to be a non-countable classifier that must pluralise to combine with numerals (including as last resort in Welsh), while the packaging singulative is taken to be a portioning device.

In §6.3, it is proposed that the semantics of plurality and individuation are compatible with a split DP approach to nominal number, where the functions of number proposed in §6.2 are stackable. In this section, I propose that an Economy constraint rules out the Welsh plural of the singulative on vacuous grounds while legally deriving the Hejazi/Modern Standard Arabic plural of the singulative.

In §6.4, I offer a brief discussion of the consequences for the theory in terms of the interpretation of inclusive and exclusive plurality.

6.1 Setting the Scene

6.1.1 The puzzle

Let us briefly recap some of the key data and conclusions that present a puzzle for a theory of number.[1] First, Welsh and Arabic have productive plural markers, where the plural is the morphologically marked form.

(1) Gwelodd Owain **ferch-ed** yn yr ardd — Welsh
see.PST.3 Owain girl-PL in DEF garden
'*Owain saw girls in the garden*'

(2) Malik shaaf **ban-aat** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 girl-PL in DEF-garden
'*Malik saw girls in the garden*'

Alongside the singular/plural system, Welsh and Arabic have a collective/singulative system which is a highly productive number neutral uncountable system akin to the English object mass system and the Mandarin Chinese general number system. This is shown by the fact that collectives are a.) number neutral as shown in (3), (4), and b.) cannot combine directly with numerals, and an intermediary structure (the singulative) is required for legal numeral composition as shown in (5) and (6). In this way, individuating singulative markers and classifiers are akin in function. Note, when combining with numerals, individuating singulatives follow language-specific agreement patterns in numeral contexts. The Welsh individuating singulative exhibits compulsory singular agreement in numeral contexts as in (5), while the Hejazi/Modern Standard Arabic singulative has multiple distinct agreement patterns, dependent on the presence of higher/lower numerals and and alternation with broken plural patterns. The Hejazi Arabic example in (6) experiences feminine sound plural agreement.

[1] Note, I only show Hejazi Arabic in this chapter for ease of exposition, but the facts also apply to Modern Standard Arabic, though see chapter 2 for the Hejazi/Modern Standard Arabic data.

(3) Gwelodd Owain **hwyaid** yn yr ardd — Welsh
see.pst.3 Owain duck in def garden
'*Owain saw ducks/a duck in the garden*'

(4) Malik shaaf **baṭ** fi al-hadiqa — H. Arabic
Malik see.pst.m duck in def-garden
'*Malik saw a duck / ducks in the garden*'

(5) Gwelodd Owain **bum hwyad-en** yn yr ardd — Welsh
see.pst.3 Owain five duck-sing in def garden
'*Owain saw five ducks in the garden*'

(6) Malik shaaf **khams-ah baṭ-a-at** fi al-hadiqa — H. Arabic
Malik see.pst.m.3 five-f duck-sing-pl in def-garden
'*Malik saw five ducks in the garden*'

When the individuating singulative in Welsh and Hejazi/Modern Standard Arabic appears bare, it is interpreted as a strictly singular count term, as in (7) and (8).

(7) Gwelodd Owain **hwyad-en** yn yr ardd — Welsh
see.pst.3 Owain duck-sing in def garden
'*Owain saw a duck in the garden*'

(8) Malik shaaf **baṭ-at** fi al-hadiqa — H. Arabic
Malik see.pst.m.3 duck-sing in def-garden-gen
'*Malik saw a duck in the garden*'

Being singular count, it is reasonable to expect the singulative to pluralise. This is borne out in Hejazi Arabic, for both paucal (9) and non-paucal (10) interpretations of the plural of the singulative.[2] There is, however, no plural of the singulative in Welsh. The lack of the Welsh plural of the singulative is unexpected.

[2] There is no realisation of the broken plural of the singulative for Arabic baṭ-ah.

(9) Malik shaaf **baṭ-a-at/shajar-a-at** H. Arabic
Malik see.PST.M.3 duck-SING-PAU/tree-SING-PAU
fi al-hadiqa
in DEF-garden
'*Malik saw a small number of ducks/trees in the garden*'

(10) Malik shaaf **'ashjaar** fi al-hadiqa H. Arabic
see.PST.M.3 Malik tree.SING.PL in DEF-garden
Plural reading: '*Malik saw trees in the garden*'
Abundance reading: '*Malik saw many trees in the garden*'

(11) *Gwelodd Owain **hwyad-enn-au** yn yr ardd Welsh
see.PST.3 Owain duck-SING-PL in DEF garden
Intended: '*Owain saw ducks in the garden*'

The aim of this chapter is therefore simple: I will provide a semantics that a.) aligns the primitive features of singulative/general classifier semantics across English, Mandarin Chinese, Welsh and Hejazi/Modern Standard Arabic, and b.) rules out the plural of the singulative in Welsh.

6.1.2 On plural markers

The developments of the semantics of classifiers and plural markers in this chapter will not be uncontroversial. Depending on the theory, plural markers may or may not have a unified semantics in numeral and non-numeral contexts. For some authors, plural marking has a unified semantics in numeral and non numeral contexts (Landman, 2020). For other authors, plural markers only have plural semantics when bare, and offer no semantic contribution in numeral constructions where plural marking is triggered by agreement (Deal, 2017; Ionin and Matushansky, 2004; Krifka, 1989; Wągiel, 2021b). In other theories of number, neither singular nor plural number marking offers a semantic contribution, but both are reflexes of agreement, *i.e.* there is a check-point where agreement is determined on a basis of whether the noun denotes (contextually) atomic/singular individuals or otherwise (Chierchia, 2010b, 2021; Sauerland et al., 2005; Scrontas, 2014). There are also the polysemous accounts of plural marking, where plural markers are ambiguous between multiple semantics across bare and numeral contexts (Martí, 2020; Dali and Mathieu, 2021b). Finally, there is the Borer (2005a) approach where plural markers do not indicate plurality at all, but are reflexes of classification/division.

Even when plural markers appear bare (and so cannot be agreement), there is no clear consensus regarding their semantic contribution. The `INCLUSIVE/AMBIGUITY` debate concerns the proper analysis of the interpretation of bare plural markers in different contexts. At its most intuitive level, the interpretation of a plural marked term is *two or more*. For example, the sentence in (12) is true if and only if Edward saw two or more girls in the garden. Contrastively, in downward entailing contexts and questions, the interpretation of the plural marked term is *one or more*: examples (13)-(15) are false even if Edward saw a single girl in the garden.

(12) Edward saw **girls** in the garden — English

(13) Edward did not see **girls** in the garden — English

(14) If Edward sees **girls** in the garden, he will inform the authorities — English

(15) Q: Did Edward see **girls** in the garden? — English
A: Yes, he saw **one**.

The controversy concerns whether bare plural markers have a unified or ambiguous semantics in these contexts. Ivlieva (2013); Mayr (2015); Sauerland et al. (2005); Spector (2007); Zweig (2009) adopt the `INCLUSIVE APPROACH` to the bare plural where the semantics of the plural marker is taken to be inherently inclusive, and therefore include reference to the singular, *i.e.* there is reference to disjoint elements and their sums. Any interpretation where plural marked terms do not refer to singulars are derived via (scalar) implicature. The general picture of the inclusive approach to plural markers is sketched out in (16).

(16) a. Edward saw girls in the garden

Literal meaning: '*Edward saw one, or more than one, girl in the garden*'

Derived meaning: '*Edward saw two or more girls in the garden*'

b. Edward did not see girls in the garden

Literal meaning: '*Edward did not see one, or more than one, girl in the garden*'

No derived meaning. Literal meaning arises in context.

On the other hand, Borer (2005a); Borer and Ouwayda (2021); Dali and Mathieu (2021b); Farkas and de Swart (2010); Martí (2017, 2020) adopt the `AMBIGUITY APPROACH` to bare plural markers, where plural marking is ambiguous between (at least) two meanings and/or structures. The first meaning is the inclusive meaning, where plural semantics includes reference to the singular. The second meaning of plural marked terms is exclusive, where the plural does not include reference to singularities, *i.e.* mereological sums only. For these accounts, the choice of interpretation is derived via pragmatic reasoning in context, such as Heim (1991)'s `MAXIMIZE PRESUPPOSITION`. The general picture of the ambiguity approach to plural markers is mapped out in (17).

(17) a. Edward saw girls in the garden

Meaning one: '*Edward saw one, or more than one, girl in the garden*'

Meaning two: '*Edward saw two or more girls in the garden*'

Choice: Meaning two is chosen in context

b. Edward didn't see girls in the garden

Meaning one: '*Edward did not see one, or more than one, girl in the garden*'

Meaning two: '*Edward did not see two or more girls in the garden*'

Choice: Meaning one is chosen in context

Clearly, the issues surrounding plural semantics and plural interpretation in context is complex. In what follows, I will be following Landman (2020)'s Iceberg Semantics and de Vries and Tsoulas (2021)'s Subpredicative Iceberg Semantics (implicit) approach to plural markers. I take plural markers as the hallmark of identity, *i.e.* plural markers are a reflex of semantic plurality when bare or otherwise. Second, I will assume that numerals must combine with semantically plural nominal terms, specifically ⟨***body**, **base**⟩ terms where the **base** is disjoint and the **body** is closed under sum. As we will see below, this requires an assumption that the phonological realisation of plural markers is contextually licensed. Finally, I will be assuming an *inclusive* approach to plural marking, where any exclusive interpretations are derived via (scalar) implicature. As it will transpire throughout this chapter, under Subpredicative Iceberg Semantics assumptions the inclusive approach to plurality is compatible with the fact that the Welsh plural of the singulative is absent, but the exclusive approach is not readily compatible with this fact.

6.2 Basic functions of number

6.2.1 Preliminaries

The analysis I put forward is in the style of de Vries and Tsoulas (2021)'s Subpredicative Iceberg Semantics model developed in chapter 5. To recap, in Subpredicative Iceberg Semantics, nominal terms denote i-sets in the Landman (2020) sense, where an i-set is a pair X of **body** and **base**, formalised as ⟨body, base⟩. Both the **body** and the **base** are sets of mereological sums which are subsets of B, and the **body** generates the **base** under Link (1983)'s * (algebraic closure).

(18) `ALGEBRAIC CLOSURE`
$*P \overset{\text{def}}{=} \{x\text{: } \exists P' \ P' \neq \varnothing \wedge P' \subseteq P \wedge x = \sqcup P'\}$
The set that contains all sums of things in P
From Champollion and Krifka (2016), adapted

(19) Let B be a complete Boolean algebra.
➢ An i-set is a pair X = $\langle$**body(X), base(X)**$\rangle$ where:
body(X) $\subseteq$ B and **base(X)** $\subseteq$ B and **body(X)** $\subseteq$ ***base**(X)
and $\sqcup$**body(X)** = $\sqcup$**base(X)**.
From Landman (2020, p.143)

In the Iceberg Semantics fashion, the disjointedness status of the **body** of an i-set determines whether a nominal term is singular or plural: a disjoint **body** ensures singularity and an overlapping **body** ensures plurality as summarised in (20). On the other hand, the disjointedness status **base** of an i-set determines countability: a disjoint **base** ensures countability and an overlapping **base** ensures uncountability as summarised in (21). With these preliminaries set, the basic system of nominal number has four basic nominal types as mapped out in (22).

(20) Let X be an i-set
➢ X is *singular* iff **body**(X) is disjoint
➢ X is *non-singular* iff X is non-null then X is not singular.

(21) Let X be an i-set
➢ X is *count* iff **base**(X) is disjoint
➢ X is *mass* iff X is non-null then X is not count.
From Landman (2020, p.165), adapted

(22) NOMINAL TYPES IN SUBPREDICATIVE ICEBERG SEMANTICS

Singular count	$\langle$**body**, **base**$\rangle$	
Number neutral countable	$\langle$***body**, **base**$\rangle$	
Number neutral uncountable	$\langle$***body**, ***base**$\rangle$	
Stuff mass	$\langle$**body**, **base**$\rangle$	(where **base** is not disjoint)

I follow de Vries and Tsoulas (2021)'s adaptation of Iceberg Semantics such that i-sets are compositionally derived. The derivations for a singular count i-set is in Figure 6.1, and the derivation for a number neutral uncountable (including collectives and general number) i-set is in Figure 6.2. The minimal difference in composition is the projection of pluralisation of the filtered set: in number neutral uncountable case, the filtered set is closed under Linkian sum, ensuring that both the **body** and the **base** of the resulting i-set is overlapping.[3]

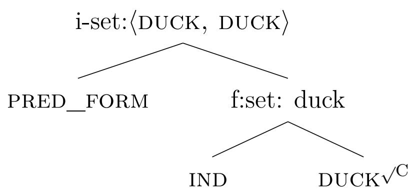


Figure 6.1: de Vries and Tsoulas' structure for singular count terms

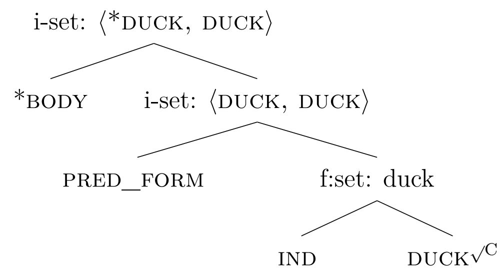


Figure 6.2: de Vries and Tsoulas' structure for number neutral countable terms

Number semantics in Subpredicative Iceberg Semantics is split across the **body** or the **base** of an i-set. Therefore, any further functions of number (pluralisation, classification *etc.*) must apply to either the **body**, the **base**, or both. The basic functions in (23) and (24) are assumed so the semantics has access to the **body** and **base**, respectively.

(23) BODY(P) is a function that maps any i-set P to the set of its **body**.

(24) BASE(P) is a function that maps any i-set P to the set of its **base**.

With these preliminaries set, I now turn to offer a semantics of plurality and classification in Welsh and Arabic, with the ultimate aim to build a system that predicts that the singulative pluralises in Arabic, but not Welsh.

[3]See chapter 5 for details on the projection of **f:set* across languages.

6.2.2 Countable plurals

For an i-set to be countable, it must have a disjoint **base**. The idea is that the **base** supplies, in context, the terms in which the **body** is counted (Landman, 2020, p.152). The **base** must supply a disjoint standard for operations that presuppose disjointedness. In this way, any i-set which combines with a numeral (higher than *one*) must be $\langle$***body**, **base**$\rangle$. The overlapping **body** ensures plurality for numeral intersection, and the disjoint **base** supplies, in context, the disjoint standard in which the **body** is counted. As such, plural markers in numeral constructions are reflexes of a semantic operation akin to the schematic in (26).

(25) Schematic for countable plural semantics
$\textsc{pl}_{\textsc{c}}(\mathrm{P}) = \langle \mathbf{body}, \mathbf{base} \rangle \rightarrow \langle {}^*\mathbf{body}, \mathbf{base} \rangle$

A formal semantics for countable plurals is simple, and is defined in (26).[4] $\textsc{pl}_{\textsc{c}}(\mathrm{P})$ applies to a singular count i-set $\langle$**body**, **base**$\rangle$ which has a disjoint **base**, extracts the **base**, and uses it to generate the **body** and the **base** of the resulting i-set. The **body** of the new i-set is closed under Linkian sum. Note, the input for the $\textsc{pl}_{\textsc{c}}(\mathrm{P})$

[4] Landman (2020) offers the semantics in (i) for the semantics of English plural -S, based on a general HEAD PRINCIPLE for NPs in (ii). The argument Landman puts forward in favour of the head principle is to ensure that **base** information is passed up from the interpretation of the head of a complex NP.

(i) $\mathrm{PL} = \lambda P.\langle {}^*body(p), ({}^*body(p)] \cap base(p)\rangle$

(ii) Head Principle for complex NPs:
$\mathbf{base}(\alpha) = (\mathbf{body}(\alpha)] \cap \mathbf{base}(\mathrm{H}_\alpha)$
The **base** of a complex NP is the intersection of the part set of the **body** of that NP and the **base** of the head NP.
From (Landman, 2020, p. 146)

While (i) works for countable plurals, I do not adopt Landman (2020)'s semantics for plural terms nor The Head condition. The reasoning is that a semantics for uncountable plurals is incompatible with The Head Principle. A minimal adaption from (i) for uncountable plurals is in (iii). In these semantics, the **base** of the complex uncountable plural is closed under sum, ensuring uncountability. This violates Landman's head principle as the **base** of the resulting uncountable plural NP is not the intersection of the Boolean part set of **body** of the uncountable plural NP and the **base** of the head. There is no way to rectify this in such a way that captures both the semantics of uncountable plurals and respects Landman's head principle.

(iii) Uncountable $\mathrm{PL} = \lambda P.\langle {}^*body(p), {}^*(({}^*body(p)] \cap base(p))\rangle$

is presupposed to be an i-set with a disjoint **base**, ruling out composition with stuff mass and number neutral uncountable i-sets.

An explicit derivation of a plural count term is given in (27), where assuming English CAT denotes a singular count ⟨**body**, **base**⟩ i-set, the output is a ⟨***body**, **base**⟩ i-set realised as CATS. $\text{PL}_{\text{C}}(\text{P})$ is assumed to project directly above a ⟨**body**, **base**⟩ i-set as in Figure 6.3, which is filled with phonological content from English for ease of exposition.

(26) $$\text{PL}_{\text{C}}(\text{P}) \rightarrow \lambda \text{P}. \begin{cases} \langle {}^{*}\text{BASE}(\text{P}),\ \text{BASE}(\text{P}) \rangle & \text{if BASE(P) is disjoint} \\ \perp & \text{otherwise} \end{cases}$$

(27) GIRL-S (*'girl-pl'*)

$= \text{PL}_{\text{C}}(\text{P})(\text{GIRL})$

$= \lambda \text{P}. \langle {}^{*}\text{BASE}(\text{P}),\ \text{BASE}(\text{P}) \rangle\ (\langle \text{GIRL},\ \text{GIRL} \rangle)$

$= \langle {}^{*}\text{BASE}_{\langle \text{GIRL, GIRL} \rangle},\ \text{BASE}_{\langle \text{GIRL, GIRL} \rangle} \rangle$

$= \langle {}^{*}\text{GIRL},\ \text{GIRL} \rangle$

☞ GIRLS is a plural count i-set

It is plural, because $\text{BODY}_{(\text{GIRLS})}$ is closed under *

It is countable, because $\text{BASE}_{(\text{GIRLS})}$ is disjoint

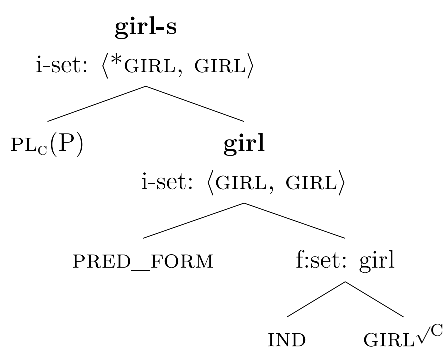


Figure 6.3: Countable plural i-set composition

Moving to our languages of interest, Welsh and Hejazi Arabic, the Subpredicative Iceberg Semantics model easily accounts for the semantics of countable plurals, though spell-out is complicated. In Arabic, plural marked terms are countable as in the feminine sound plural in (28). In this case, feminine sound plural marking can be assumed to be a reflex of the PL$_{C}$(P) function which derives the $\langle$*$\mathbf{body}$, $\mathbf{base}\rangle$ i-set. The Arabic case is notable in so much as there is complexity in spell-out, including affixation in the case of a sound plural (which agrees in gender) as shown in (28), but also stem alteration in the case of a broken plural, and zero spell out when combining with numerals higher than ten (see §2.2.2 for more examples of Hejazi/Modern Standard Arabic plural marking in different contexts).

(28) Malik shaaf **khams-ah ban-aat** fi al-hadiqa — H. Arabic
see.PST.M.3 Malik five-F girl-PL in DEF-garden
'*Malik saw five girls in the garden*'

The case of the Welsh is a different matter. While Welsh has a plural marker as in (29), Welsh countable nominals are always formally singular-marked in numeral contexts. In the case of singular count terms, this is zero-marked, as in (30). The immediate analysis is that Welsh has a covert PL$_{C}$(P) projection in numeral contexts.

(29) ferch → ferch-ed — Welsh
GIRL (girl.SG = *'girl'*) → GIRL-S (girl.SG-PL = *'girls'*)

(30) Gwelodd Owain **bum merch** yn yr ardd. — Welsh
see.PST.3 Owain five girl in DEF garden
'*Owain saw five girls in the garden*'

A covert analysis of Welsh PL$_{C}$(P) is necessitated by the theory, but there is independent evidence to support this claim. Consider the Welsh split demonstrative which agrees with the head noun in number, where HON (*'this'*) is singular as in (31) and HYN (*'these'*) is plural as in (32). In the case that demonstratives and numerals are used in combination, the noun appears formally unmarked, yet the agreement for the demonstrative is plural, as in (33).[5] I suggest that these are simple agreement facts: the noun AFAL in (33) is plural, but pluralisation is realised covertly in the context of numerals, and the demonstrative HYN agrees accordingly.[6]

(31) Prynodd Owain **yr** **afal** **hwn** ddoe — Welsh
buy.PST.3 Owain DEF apple DEM.SG.M yesterday
Owain bought this apple yesterday

(32) Prynodd Owain **yr** **afal-au** **hyn** ddoe — Welsh
buy.PST.3 Owain DEF apple-PL DEM.PL yesterday
Owain bought these apples yesterday

(33) Prynodd Owain **y** **pum** **afal** **hyn** ddoe — Welsh
buy.PST.3 Owain DEF five apple DEM.PL yesterday
Owain bought these five apples yesterday

[5]For completion, note that collective/singulative terms in these contexts exhibit a parallel pattern: where the zero-marked singular appears, the singulative appears, and where the plural marked term appears, the collective appears. This is unsurprising, as singulatives and singulars are semantically akin, and as will be discussed in §6.2.3 collectives are semantically akin to Welsh plural marked terms. Note, the singular demonstrative agrees in gender, and orthography facts follow from mutation facts (see §1.5.2).

(i) Prynodd Owain **yr** **ellyg-en** **hon** ddoe — Welsh
buy.PST.3 Owain DEF pear-SING DEM.SG.F yesterday
Owain bought this pear yesterday

(ii) Prynodd Owain **y** **gellyg** **hyn** ddoe — Welsh
buy.PST.3 Owain DEF pear DEM.PL yesterday
Owain bought these pears yesterday

(iii) Prynodd Owain **y** **pum** **gellyg-en** **hyn** ddoe — Welsh
buy.PST.3 Owain DEF five pear-SING DEM.PL yesterday
Owain bought these five pears yesterday

[6]In 6.2.3, I offer a semantics for the Welsh overt plural that minimally differs from the covert plural. The plural semantics for overt and covert plurals are united by an overlapping **body** (*i.e.* closure under sum). I take the demonstrative agreement in Welsh to directly result from agreement with the overlapping **body** of plural terms, overt, covert, countable or uncountable. An alternate approach is by Ionin and Matushansky (2006) who argue that numerals combine with semantically singular nouns and plural marking in *e.g.* English is not semantic plurality but semantic concord. In this manner, Welsh non-marking in numeral contexts is the 'truth' of number marking where atomic/singular nouns are required for numeral construction.

6.2.3 Uncountable plurals

I follow both Landman (2020) and de Vries and Tsoulas (2021) in taking $\langle {}^{*}\mathbf{body}, {}^{*}\mathbf{base}\rangle$ i-sets as the basic nominal type of number neutral uncountable terms where the closure of the **body** under * ensures plurality, and the closure of the **base** under * ensures uncountability. Expanding on this, I suggest that the Welsh overt plural is a function which derives a $\langle {}^{*}\mathbf{body}, {}^{*}\mathbf{base}\rangle$ term in the manner of the schematic in (34). This is because the Welsh overt plural may appear bare, as in repeated example (1), but is outright ungrammatical when directly combining with numerals (35).

(34) Schematic for derived uncountable plural semantics
$\textsc{pl}_{\textsc{nc}}(\mathrm{P}) = \langle \mathbf{body}, \mathbf{base}\rangle \rightarrow \langle {}^{*}\mathbf{body}, {}^{*}\mathbf{base}\rangle$

(1) Gwelodd Owain **ferch-ed** yn yr ardd — Welsh
see.PST.3 Owain girl-PL in DEF garden
'*Owain saw girls in the garden*'

(35) *Gwelodd Owain **pum ferch-ed** yn yr ardd. — Welsh
see.PST.3 Owain five girl-PL in DEF garden
'*Owain saw five girls in the garden*'

A formal account of the schematic in (34) is in (36). $\textsc{pl}_{\textsc{nc}}(\mathrm{P})$ applies to a singular count i-set $\langle \mathbf{body}, \mathbf{base}\rangle$ which has a disjoint **base**, extract the **base**, and uses it to generate the resulting i-set. Both the **body** and the **base** of the new i-set are closed under sum, ensuring both plurality and uncountability. A derivation for Welsh CATHOD (cat.SG-PL$_{\textsc{nc}}$ = *'cats'*) is given in (37), with a structural representation in Figure 6.4, where $\textsc{pl}_{\textsc{nc}}(\mathrm{P})$ is taken to project above a $\langle \mathbf{body}, \mathbf{base}\rangle$ i-set. The derivation in (37) is filled with phonological content from Welsh for ease of exposition.

(36) $\textsc{pl}_{\textsc{nc}}(\mathrm{P}) \rightarrow \lambda \mathrm{P}.\begin{cases} \langle {}^*\textsc{base}(\mathrm{P}),\ {}^*\textsc{base}(\mathrm{P})\rangle & \text{if } \textsc{base}(\mathrm{P}) \text{ is disjoint} \\ \perp & \text{otherwise} \end{cases}$

(37) $\textsc{cath-od}$ (cat.$_{\textsc{sg-pl}_{\textsc{nc}}}$ = *'cats'*)
$= \textsc{pl}_{\textsc{nc}}(\mathrm{P})(\textsc{cath})$
$= \lambda \mathrm{P}.\langle {}^*\textsc{base}(\mathrm{P}),\ {}^*\textsc{base}(\mathrm{P})\rangle\ (\langle \textsc{cath},\ \textsc{cath}\rangle)$
$= \langle {}^*\mathrm{BASE}_{\langle \textsc{cath},\ \textsc{cath}\rangle},\ {}^*\mathrm{BASE}_{\langle \textsc{cath},\ \textsc{cath}\rangle}\rangle$
$= \langle {}^*\textsc{cath},\ {}^*\textsc{cath}\rangle$

☞ CATHOD is a plural non-count i-set

It is plural, because $\textsc{body}_{(\textsc{cathod})}$ is closed under *

It is non-countable, because $\textsc{base}_{(\textsc{cathod})}$ is not disjoint

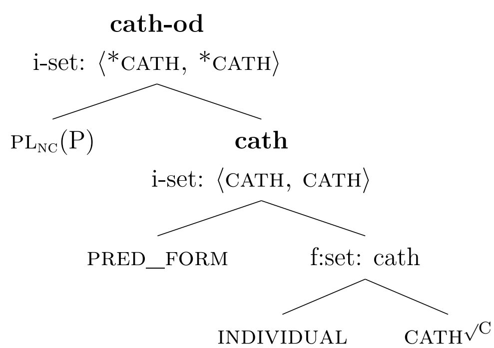


Figure 6.4: Uncountable plural i-set composition

The semantics for the Welsh overt plural derives a term with identical number semantics to a collective, *i.e.* they are both ⟨***body**, ***base**⟩ i-sets. The only difference between Welsh collective and Welsh overt plurals is in the manner of composition as either subpredicative (as in collectives) or predicative (as in uncountable plurals) summarised in (38).

(38) ⟨***body**, ***base**⟩ i-sets in Welsh:

Number neutral uncountable (incl. collectives)
derived via **f:set* at the subpredicative level
e.g. HWYAID (duck.COL = *'duck(s)'*)

Uncountable Plurals
derived via PL_{NC}(P) at the predicative level
e.g. CATH-OD (cat.SG-PL_{NC} = *'cats'*)

Treating the Welsh overt plural and and collective terms as semantically akin (number neutral uncountable) has independent support beyond the fact that neither can combine directly with numerals. That is, they may both be targetted by the same pseudopartitive structure. As will be seen in §6.3.1, I take the pseudopartitive construction to apply only to predicates of ⟨***body**, ***base**⟩ status.

(39) Gwelodd Owain **bump o hwyaid** yn yr ardd — Welsh
see.PST.3 Owain five of duck in DEF garden
'*Owain saw five ducks in the garden*'

(40) Gwelodd Owain **bump o ferch-ed** yn yr ardd — Welsh
see.PST.3 Owain five of girl-PL in DEF garden
'*Owain saw five girls in the garden*'

In a sense this analysis makes the Welsh overt plural a 'derived collective'. This should not be confused with the 'derived collective' in Czech and Ukrainian. According to Grimm and Dočekal (2021) for Czech and Wągiel and Shlikhutka (2023a) for Ukrainian, the collective deriving morphemes in table 6.1 induce a sense of clustering. This is fundamentally different from the Welsh overt plural which has no sense of clustering. I use the term 'derived collective' as I assume the same primitive components ⟨***body**, ***base**⟩ underlie both the overt plural and the collective.[7]

[7] Grimm and Dočekal and Wągiel and Shlikhutka report that Ukrainian/Czech derived collectives are uncountable. It very well be the case that a ⟨***body**, ***base**⟩ analysis for the underlying semantics may be appropriate here, though the sense of clustivity would need to be incorporated.

Language	Singular	Derived Collective	Gloss
Czech	strom	strom-oví	tree → clump of trees
	list	listí	leaf → foliage
Ukrainian	pero	pir-j-a	feather→ clump of feathers
	lyst	lyst-j-a	leaf → foliage

Czech examples are from Grimm and Dočekal (2021, p.92)
Ukrainian Examples are from Wągiel and Shlikhutka (2023a)

Table 6.1: Ukrainian and Czech derived collectives

6.2.4 Excursus: on covertness, overtness, and last resort

I have suggested that Welsh has both a covert $\textsc{pl}_{\textsc{c}}$(P) that is used in the context of numerals, and an overt $\textsc{pl}_{\textsc{nc}}$(P) that is used otherwise. An immediate question arises: why is the Welsh $\textsc{pl}_{\textsc{c}}$(P) restricted to numeral environments? After all, $\textsc{pl}_{\textsc{c}}$(P) is not restricted to numeral environments in English nor Arabic.

I suggest that the Welsh overt $\textsc{pl}_{\textsc{nc}}$(P) is the primary (and free) plural marker, and covert $\textsc{pl}_{\textsc{c}}$(P) is a last resort device induced only when $\textsc{pl}_{\textsc{nc}}$(P) fails, *i.e.* in the context of numerals. To see why, consider that covert $\textsc{pl}_{\textsc{c}}$(P) and overt $\textsc{pl}_{\textsc{nc}}$(P) output i-sets which are identical in the **body** and differ only in the disjointedness of the **base**, as summarised in (41). Now, the status of the **base** is only relevant in counting contexts: a disjoint **base** is required for combination with numerals. In numeral free contexts interpretation of an i-set takes place on the **body**. In this way, $\textsc{pl}_{\textsc{nc}}$(P) and $\textsc{pl}_{\textsc{c}}$(P) are interpreted identically in numeral free contexts.

(41) $\textsc{pl}_{\textsc{c}}(\text{P}) = \langle \textbf{body}, \textbf{base} \rangle \rightarrow \langle {}^*\textbf{body}, \textbf{base} \rangle$ (Welsh covert plural)
$\textsc{pl}_{\textsc{nc}}(\text{P}) = \langle \textbf{body}, \textbf{base} \rangle \rightarrow \langle {}^*\textbf{body}, {}^*\textbf{base} \rangle$ (Welsh overt plural)

To modulate the choice of Welsh PL_{C}(P) and PL_{NC}(P) in numeral free contexts I quote Chierchia who puts it succinctly:

> Don't do covertly what you can do overtly.
> (Chierchia, 1998a, p.360)

Through general reasoning, unmarked terms in numeral free contexts are never interpreted as covertly pluralised via PL_{C}(P), even if they theoretically could be: if plural meaning was intended, then the overt form PL_{NC}(P) would have been used. In numeral contexts, the overt plural PL_{NC}(P) is ungrammatical and can never arise due to the overlapping **base**, and therefore covert PL_{C}(P) must be used. In this way, covert pluralisation in Welsh is only predicted in numeral contexts as a last resort mechanism.

There is a crosslinguistic parallel with regards to Turkish at this point. Like Welsh, Turkish has an overt plural, -LAR[8], as in (42), which is uncountable (43), and instead, the unmarked form of the noun is used in numeral constructions (44). However, Turkish differs from Welsh in that unmarked nominals in object position are interpreted as number neutral as in (45).

(42) Ömer bahçe-de **kiz-lar** gördü — Turkish
Ömer garden-in girl-PL see.PST.3
'*Ömer saw girls (2+) in the garden*'

(43) *Ömer bahçe-de **beş kiz-lar** gördü — Turkish
Ömer garden-in five girl-PL see.PST.3
'*Ömer saw five girls in the garden*'

[8] -LAR marked plurals (42) have been argued to have a strong preference for exclusive interpretations even in downward entailing contexts, and so have been argued to be inherently exclusive (Bliss, 2004; Martí, 2020). However, *c.f.* Renans et al. (2020) for experimental evidence to the contrary.

(44) Ömer bahçe-de beş **kiz** gördü — Turkish
Ömer garden-in five girl see.PST.3
'*Ömer saw five girls in the garden*'

(45) Ömer bahçe-de **kiz** gördü — Turkish
Ömer garden-in girl see.PST.3
'*Ömer saw a girl / girls in the garden*'

To account for examples like (44) and (45), de Vries and Tsoulas (2021) posit that a PL_C(P) projection is compulsory and covert for every i-set derivation in Turkish. The idea is that bare nominals in Turkish are default unmarked ⟨***body**, **base**⟩ (number neutral countable) terms, and so always interpreted as number neutral when bare, as in (45), but are also compatible with numerals without further pluralisation functions, as in (44). This cannot be correct. As pointed out by Sağ (2022) it is not the case that bare nominals are always interpreted as number neutral. For example, ÇOCUK (*'child'*) in (46) is undefined if more than one child ran home.

(46) **Çocuk** ev-e koş-tu — Turkish
child home-DAT run-PST
'*The child ran home*'
From (Sağ, 2022, p.744)

Sağ (2022) suggests that unmarked terms in Turkish are singular count terms, like in English, and number neutral interpretations in examples such as (45) arise due to a process of incorporation. This view is compatible with the version of Subpredicative Iceberg Semantics set out here. That is, Turkish nominals are exactly like Welsh nominals: by default, they are ⟨**body**, **base**⟩ i-sets (singular count), and have an overt PL_NC(P) function (-LAR) which is semantically incompatible with numerals, and therefore a covert (last resort) operation must apply in numeral contexts. Turkish therefore differs from Welsh in that Turkish nominals undergo a process of pseudo-incorporation for bare nominals in *e.g.* object position as in (45).

With this, there is a question regarding the crosslinguistic overt/covert nature of countable plurals, *i.e.* PL_{C}(P). I tentatively suggest that when a language has an overt plural that is uncountable, *i.e.* PL_{NC}(P), then PL_{C}(P) arises as a covert last resort mechanism in numeral constructions. The data here suggest this is the case in Welsh and Turkish (and as we will see, Mandarin), but a much larger sample of languages is needed to test whether this suggestion holds.[9]

***End of excursus**

6.2.5 General classifiers and individuating singulatives

The Mandarin Chinese general classifier and the Welsh and Arabic individuating singulatives are semantically akin: they combine with a number neutral uncountable term and output a semantic singular, imparting no semantics other than individuation. In this manner, I follow the standard analysis that singulatives and general classifiers share the same primitive semantic functions (Mathieu, 2012a, 2013, 2014; Fassi Fehri, 2004, 2012, 2016, 2020; Zabbal, 2002; Jaradat and Jarrah, 2022).

Landman (2020, p.212) already proposes a semantics for the Mandarin Chinese general classifier, GÈ, as a function that maps an i-set with an overlapping **body** and **base** onto an i-set with an overlapping **body** and a disjoint **base**, as schematised in (47).[10] The reasoning is that classifiers in Mandarin Chinese combine with numerals, so the disjoint **base** ensures countability while the overlapping **body** ensures that intersection with numerals is non-empty.

[9] There is a reasonable question regarding whether there exists languages where the opposite state of affairs hold, *i.e.* where a language has a covert countable plural, PL_{NC}(P), but an overt countable plural, PL_{C}(P). There is a second reasonable question as to whether there exists languages where both types of plurals are overt. I am unsure whether these types of language exist and offer no analysis for these possibilities, assuming that covertness in plural contexts arises due to last resort.

[10] Landman (2020, p.320) actually distinguishes between countable and non countable classifiers. For Landman a semantics like (47) is not appropriate for classifiers in languages that make number distinctions, such as English and Dutch. Instead, for these languages Landman argues for a semantics more in line with (49). In a strict adherence to Landman's Iceberg Semantics, we may analyse individuating singulatives and general classifiers as these two separate functions, where singulatives parallel English and Dutch, while Mandarin Chinese classifiers are a different type.

(47) Schematic for Mandarin Chinese classifier semantics as in Landman (2020):
$\langle {}^*\textbf{body}, {}^*\textbf{base}\rangle \rightarrow \langle {}^*\textbf{body}, \textbf{base}\rangle$

The semantics Landman proposes for GÈ are inappropriate to import into the semantics of singulative markers. The overlapping **body** in the output of the function in (47) ensures plurality of the resulting i-set, but the singulative marked terms are strictly singular: examples (7) and (8), repeated here for convenience, have no plural interpretation. In fact, it is not clear that the schematic in (47) is an appropriate semantics for Mandarin Chinese either, as the bare classifier in (48) is interpreted as singular. A more appropriate semantics that unifies both individuating singulatives and generalised classifiers would be where a singular count term is derived from a number neutral uncountable as schematised in (49).

(7) Gwelodd Owain **hwyad-en** yn yr ardd — Welsh
see.PST.3 Owain duck-SING in DEF garden
'*Owain saw a duck in the garden*'

(8) Malik shaaf **baṭ-at** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 duck-SING in DEF-garden-GEN
'*Malik saw a duck in the garden*'

(48) Shīyáng zài huāyuán lǐ kàndào-le **gè nǚhái** — M. Chinese
Shīyáng at garden in see-PST CL girl
'*Shīyáng saw a girl in the garden.*'

(49) Schematic for singular count deriving classifier semantics
$\text{CL}_{\text{SING}}(\text{P}) = \langle {}^*\textbf{body}, {}^*\textbf{base}\rangle \rightarrow \langle \textbf{body}, \textbf{base}\rangle$

Now we formalise the schematic in (49). First, we need to access the disjoint individuals which comprise the **base** of the $\langle {}^*\textbf{body}, {}^*\textbf{base}\rangle$ input. To do this, a few ingredients are needed, the first being `BOOLEAN PART SET` as in (50). A Boolean part set is a function which takes a set or a sum as domain and outputs all parts, which in itself creates a mereology, as range. Following Landman (2020), who follows Grätzer (1998), I use half-closed interval notation, $(x]$, to represent Boolean part sets.[11] An example of a Boolean part set operation is as in (51), where the part set of the join $a \sqcup b \sqcup c$ is the set that contains $a \sqcup b \sqcup c$ and all its parts.

(50) Let $x \in B$, $X \subset B$
Boolean part set of x, $(x]$, is given by: $(x] = \{b \in B\text{: } b \sqsubseteq x\}$
Boolean part set of X, $(X]$, is given by: $(X] = (\sqcup X]$
From (Landman, 2020, p.20)

(51) $(a \sqcup b \sqcup c] = \{\mathrm{a, b, c}, a \sqcup b, a \sqcup c, b \sqcup c, a \sqcup b \sqcup c\}$

I now define IPARTS(b) in (53), which maps a set or sum *b* to the intersection of that set *b* and the set of pre-theoretical individuals (see §3.5 for discussion of pre-theoretical individuals and the reasoning behind (52)).

(52) $\textsc{individual}(x) \overset{\text{def}}{=}$
$\forall P[P(x) \rightarrow (\textsc{mssc}(x,P) \vee \textsc{scatter}(x,P) \ldots)]$
x is an individual if. for any property P x is maximally strongly self connected relative to property P, or is a scattered whole relative to property P ...

(53) $\mathrm{IP}\textsc{arts}(b) =$
$\lambda b.[(b] \cap \{x\text{: } \textsc{individual}(x)\}]$

I use IPARTS(b) as the main semantic component of $\textsc{cl}_{\textsc{sing}}(P)$ in (54), the unified function for the individuating singulative and the Mandarin Chinese generalised classifier. $\textsc{cl}_{\textsc{sing}}(P)$ applies to a $\langle$***body**, ***base**$\rangle$ predicate, extracts its **base**, and applies it to the IPARTS(b), for which the output serves as the disjoint **body** and **base** of the newly derived singular count term. An example $\textsc{cl}_{\textsc{sing}}(P)$ derivation

[11] Landman cites Grätzer's first edition rather than the updated 1998 version I use here.

for Arabic BAṬ-AH (duck.COL-SING = *'duck'*) is shown in (55). The structural representation is given in 6.5, which is filled with phonological content from Arabic for ease of exposition.

(54) $\text{CL}_{\text{SING}}(\text{P}) \rightarrow \lambda P. \begin{cases} \langle \text{IPARTS}_{\text{BASE(P)}},\ \text{IPARTS}_{\text{BASE(P)}} \rangle & \text{if P is } \langle {}^*\textbf{body},\ {}^*\textbf{base} \rangle \\ \perp & \text{otherwise} \end{cases}$

(55) BAṬ-AH (duck.COL-SING = *'duck'*)
$= \text{CL}_{\text{SING}}(\text{P})(\text{BAṬ})$
$= \lambda P.\langle \text{IPARTS}_{\text{BASE(P)}},\ \text{IPARTS}_{\text{BASE(P)}} \rangle\ (\langle {}^*\text{BAṬ},\ {}^*\text{BAṬ} \rangle)$
$= \langle \text{IPARTS}_{(\text{BASE}(\langle {}^*\text{BAṬ},\ {}^*\text{BAṬ}\rangle))},\ \text{IPARTS}_{(\text{BASE}(\langle {}^*\text{BAṬ},\ {}^*\text{BAṬ}\rangle))} \rangle$
$= \langle \text{IPARTS}_{({}^*\text{BAṬ})},\ \text{IPARTS}_{({}^*\text{BAṬ})} \rangle$
$= \langle \text{BAṬ},\ \text{BAṬ} \rangle$

☞ BAṬAH is a singular count i-set
It is singular, because $\text{BODY}_{(\text{BAṬAH})}$ is disjoint
It is countable, because $\text{BASE}_{(\text{BAṬAH})}$ is disjoint

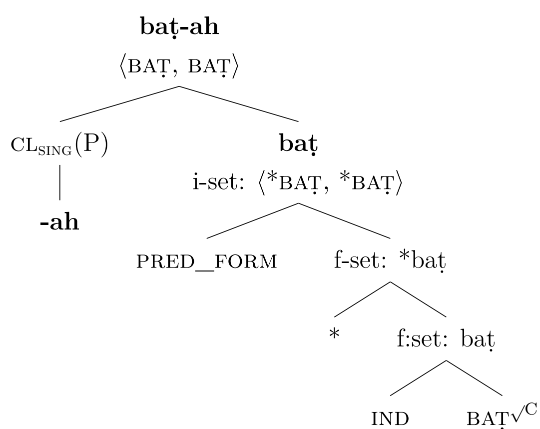


Figure 6.5: Individuating singulative and classifier i-set composition

CL$_{\text{SING}}$(P) is taken to be the unifying semantics of individuating singulatives and general classifiers. While this is the case, (surface) morphosyntactic realisations vary. I assume with Simpson (2008) that the default merge order of classifier structures is [NUM-CL-N], and so take CL$_{\text{SING}}$(P) to project directly above i-set as in Figure 6.5. Crosslinguistic variation is morphosyntactic in nature. For Mandarin Chinese GÈ, the default order is maintained. In the case of Welsh and Arabic singulative structures raising occurs (see §2.2.5 for further discussion).

6.2.6 Packaging singulatives

While the semantics of general classifiers and individuating singulatives presupposes disjoint individuals within the **base** of the i-set they combine with, the Welsh packaging singulative in (56) and its alternate stand-alone morph DARN in (57) does not.[12] At an intuitive level, the packaging singulative and DARN (*'piece'*) create individuals from mass in context. As such, the semantics of CL$_{\text{SING}}$(P) cannot be applied in these cases, as the stuff mass i-set is an incompatible input.

(56) Bwytodd Owain **gos-yn** ddoe — Welsh
eat.PST.3 Owain cheese-PSING yesterday
'*Owain ate a piece of cheese yesterday*'

(57) Bwytodd Owain **ddarn o gaws** ddoe — Welsh
eat.PST.3 Owain piece of cheese yesterday
'*Owain ate a piece of cheese yesterday*'

Following general wisdom, I use a `PARTITION` to capture the semantics for the creation of portions in context (Chierchia, 2010b; Landman, 2020; Scrontas, 2014, *a.o.*). I borrow Scrontas's definition of a partition in (58).[13] In words, a partition π of a set X is any non-empty disjoint subset of X. Take for example the mereology

[12] I focus on the Welsh packaging singulative as the Hejazi/Modern Standard Arabic is variably productive (see §2.1.2 for more details).

[13] An alternative definition of Partition with relation to countability literature can be found in Chierchia (2010b, p.21) and Landman (2020, p.22).

X in (59). Legal partitions of X include Y and Z as they are a.) subsets of X, and b.) all entities are disjoint. On the other hand, W is not a partition of X, as the elements are not disjoint ($a \sqsubseteq a \sqcup b \wedge a \sqsubseteq a \sqcup c \sqcup d$).

(58) PARTITION
π is a function of type $\langle\langle e,t\rangle,\langle e,t\rangle\rangle$
such that for any $P_{\langle e,t\rangle}$ and any x and y in π(P),
$\neg\exists$x[z $\sqsubseteq$ x $\wedge$ z $\sqsubseteq$ y]
From Scrontas (2014, p.92), adapted

(59) X = $\{a, b, c,\ a \sqcup b,\ a \sqcup c,\ b \sqcup c,\ a \sqcup b \sqcup c\}$
Y = πX = {a, d, $b \sqcup c$}
Z = πX = $\{a \sqcup b,\ c \sqcup d\}$
W $\neq \pi$X = $\{a \sqcup b,\ a \sqcup c \sqcup d\}$

We now incorporate π into the Subpredicative Iceberg Semantics model for the packaging singulative. PORT(P) in (60) applies to a stuff mass ⟨**body**, **base**⟩ term which has a conceptually overlapping **base**, extracts the overlapping **base**, applies it to π, for which the output serves as the disjoint **base** and **body** of a newly derived singular count i-set. An example is given in (61), where CAWS (cheese.SMASS = *'cheese'*) denotes a stuff mass i-set with an overlapping **base**, the output COSYN is a singular count ⟨**body**, **base**⟩ i-set. PORT(P) is assumed to project directly above a stuff mass ⟨**body**, **base**⟩ term, as exemplified in Figure 6.6, which is filled with phonological content from Welsh for ease of exposition. At this point, variation and phonological realisation is morphosyntactic in nature. For the Welsh packaging singulative, -YN/-EN, raising occurs, but for canonical portion terms such as DDARN (*'piece'*) the in-situ order is maintained.

(60) $\text{Portion(P)} \rightarrow \lambda\text{P}.\begin{cases} \langle \pi_{\textsc{base}(\text{P})},\ \pi_{\textsc{base}(\text{P})} \rangle & \text{if } \textsc{base}(\text{P}) \text{ is overlapping} \\ \perp & \text{otherwise} \end{cases}$

(61) COS-YN (cheese.$_{\text{SMASS-PSING}}$ = *'a piece of cheese'*)

$= \textsc{Port}(\text{P})(\textsc{caws})$

$= \lambda\text{P}.\langle \pi_{\textsc{base}(\text{P})},\ \pi_{\textsc{base}(\text{P})} \rangle\ (\langle \textsc{caws},\ \textsc{caws} \rangle)$

$= \langle \pi_{\mathbf{base}(\langle \textsc{caws},\ \textsc{caws} \rangle)},\ \pi_{\mathbf{base}(\langle \textsc{caws},\ \textsc{caws} \rangle)} \rangle$

$= \langle \pi_{\textsc{caws}},\ \pi_{\textsc{caws}} \rangle$

☞ COSYN is a singular count i-set

It is singular, because $\textsc{body}_{(\textsc{cosyn})}$ is disjoint

It is countable, because $\textsc{base}_{(\textsc{cosyn})}$ is disjoint

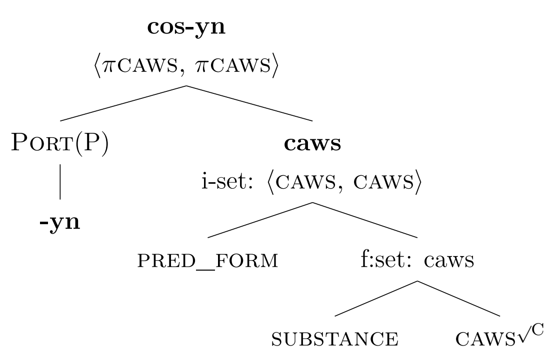


Figure 6.6: Packaging singulative i-set composition

6.2.7 Summary of the basic functions of number

Let us summarise the Subpredicative Iceberg Semantics model I have proposed so far. Following de Vries and Tsoulas (2021)'s Subpredicative Iceberg Semantics model, there are four subpredicative nominal i-set derivations, as in (22), repeated for convenience.

(22) `NOMINAL TYPES IN SUBPREDICATIVE ICEBERG SEMANTICS`

Singular count	⟨**body**, **base**⟩	
Number neutral countable	⟨***body**, **base**⟩	
Number neutral uncountable	⟨***body**, ***base**⟩	
Stuff mass	⟨**body**, **base**⟩	where the **base** is not disjoint

This chapter has introduced the functions of number in (62). These functions of number are taken to be the minimal mechanisms which derive i-sets of differing **body** and **base** properties. PL_{C}(P), PL_{NC}(P) and CL_{SING}(P) are functions which do not reconstitute the mereo(topo)logical interpretation of pre theoretical individuals denoted by an i-set. Rather, these functions of number alter the perspective of the i-set in such a way that singularity/plurality is prime (algebraic closure distinctions in the **body**) and whether the predicate can be counted (disjointedness distinctions in the **base**). Only the PORT(P) function radically relies on the alteration of the structure of the individuals denoted, where pre-theoretical substances are portioned into disjoint individuals in context.

(62) `FUNCTIONS OF NOMINAL NUMBER IN SUBPREDICATIVE ICEBERG SEMANTICS`

PL_{C}(P)	⟨**body**, **base**⟩ → ⟨***body**, **base**⟩	*e.g.* English, Arabic, Welsh covert plural
PL_{NC}(P)	⟨**body**, **base**⟩ → ⟨***body**, ***base**⟩	*e.g.* Welsh overt plural
CL_{SING}(P)	⟨***body**, ***base**⟩ → ⟨**body**, **base**⟩	*e.g.* individuating singulatives, classifiers
PORT(P)	⟨**body**, **base**⟩ → ⟨**body**, **base**⟩	*e.g.* packaging singulatives

With these minimal functions of number, the mechanisms that underlay singulativisation and classification are one and they same: they are both CL_{SING}(P) functions which derive singular count terms from number neutral uncountable terms.

(63) `UNIFIED INDIVIDUATION SEMANTICS`
Individuating singulatives and generalised classifiers are CL_{SING}(P) functions which derive singular count terms.

There is a universal claim for nominal number that follows from this model. As $\langle$*$\mathbf{body}$, $\mathbf{base}\rangle$ terms are the only i-sets which may combine with numerals above two, it must be the case that all languages have mechanisms to derive $\langle$*$\mathbf{body}$, $\mathbf{base}\rangle$ i-sets for the purpose of numeral composition. In the case of English and Arabic, the overt plural with PL_C(P) semantics serves this role; in the case of Welsh and Turkish, PL_C(P) is a covert last resort mechanism which induces PL_C(P) semantics.[14]

(64) `GENERALISATION 3: THE NUMERAL CONSTRAINT`
All languages must be able to derive $\langle$*$\mathbf{body}$, $\mathbf{base}\rangle$ terms for combination with numerals.

a. In the case that a language has an overt PL_NC(P), then PL_C(P) is induced as a (covert) last resort mechanism only in numerical contexts.

6.3 Stacking

The semantics proposed in this chapter are intentionally designed so that functions may be stacked to ultimately achieve $\langle$*$\mathbf{body}$, $\mathbf{base}\rangle$ predication which is required for combination with numerals. Therefore, functions of number are not restricted to syntactic position, nor are any functions of number in complementary distribution (*c.f.* Borer (2005a)'s account of classifiers/plural markers). In this manner, the semantics presented here are amenable to a split DP analysis of nominal number where number functions can be distributed along the nominal spine (Acuaviva, 2008; Alexiadou, 2011; Dali and Mathieu, 2021a; Mathieu, 2012a, 2013, 2014, 2012b; Wiltschko, 2008, *a.o.*). This is desirable; as singulative/classifiers are PL_C(P) functions which derive $\langle\mathbf{body}$, $\mathbf{base}\rangle$ i-sets, their output is incompatible with numerals. As such, further structure is required to derive countable i-sets.

[14]There is of course an alternative method I do not explore here: classifiers may combine directly with numerals, changing their type such that they are compatible with i-sets which are not $\langle$*$\mathbf{body}$, $\mathbf{base}\rangle$. An early proposal where classifiers combine with numerals is in Krifka (1995), but see Doetjes (2021) for general discussion.

I now turn to two areas where this split analysis of number is enlightening: the Welsh pseudopartitive, and the plural of the singulative, including a simple explanation for why there is no (overt) plural of the singulative in Welsh.

6.3.1 The Welsh pseudopartitive

The Welsh counting pseudopartitive is a curious construction. It may apply to collectives as an alternate individuation method to the singulative with little interpretative difference. The pseudopartitive may also target overtly marked plural terms as an alternative to combining numerals with an unmarked singular, again with very little interpretative difference (Borsley et al., 2007; Mittendorf and Sadler, 2005). The only difference between the pseudopartitive and other counting methods is that the pseudopartitive is strictly for counting: it cannot appear sans numeral.[15] This is shown in repeated pairs (39)/(5), and (40)/(30).[16]

(39) Gwelodd Owain *(**bump**) **o hwyaid** yn yr ardd — Welsh
see.PST.3 Owain five of duck in DEF garden
'*Owain saw five ducks in the garden*' = (5)

(5) Gwelodd Owain **(bum) hwyad-en** yn yr ardd — Welsh
see.PST.3 Owain five duck-SING in DEF garden
'*Owain saw five ducks in the garden*' = (39)

(40) Gwelodd Owain *(**bump**) **o ferch-ed** yn yr ardd — Welsh
see.PST.3 Owain five of girl-PL_{NC} in DEF garden
'*Owain saw five girls in the garden*' = (30)

[15]While there is very little interpretative difference between pseudopartitive constructions and covertly pluralised count constructions, Borsley et al. (2007) report that pseudopartitive constructions are predominantly used with higher numerals.

[16]The difference between BUM(P)/PUM(P) and MERCH/FERCH is due to phonological mutation and has no impact on the semantics. See 1.5.2 for details.

(30) Gwelodd Owain **bum merch** yn yr ardd. Welsh
see.PST.3 Owain five girl in DEF garden
'*Owain saw five girls in the garden*' = (40)

As the Welsh collective and the overt plural are both ⟨***body**, ***base**⟩ i-sets, a straightforward analysis is that ⟨***body**, ***base**⟩ i-sets serve as the input to the pseudopartitive function. The non-trivial question that follows is: what is the output of this function?

The obvious route is to treat the pseudopartitive as a realisation of $\text{CL}_{\text{SING}}(\text{P})$. This analysis has the benefit of treating the pseudopartitive and the singulatives as different realisations of the same function. At first, this seems plausible, especially as singulatives and pseudopartitives seem to have very little interpretative differences. However, this analysis leads to the postulation of redundant structure for pseudopartitives of plural marked terms as schematised in (65).[17]

(65) Pseudopartitive as a realisation of $\text{CL}_{\text{SING}}(\text{P}) + \text{PL}_{\text{C}}(\text{P})$:

MERCH	→	MERCH-ED	→	O FERCH-ED	→	∅ O FERCH-ED
girl	→	girl-PL_{NC}	→	girl-PL_{NC}-CL_{SING}	→	girl-PL_{NC}-CL_{SING}-PL_{C}
⟨**body**, **base**⟩	→	⟨***body**, ***base**⟩	→	⟨**body**, **base**⟩	→	⟨***body**, **base**⟩
				REDUNDANT		

The alternate analysis, which I advocate for, is to treat the pseudopartitive as a realisation of a plural classifier, $\text{CL}_{\text{PLUR}}(\text{P})$, as in the schematic in (66). A schematic of the derivation is given in (67). Adopting the semantics as in the schematic in (66) is beneficial not only because redundancy is ruled out, but we also have an explanation for why the Welsh pseudopartitive must always appear with a numeral: the bare ⟨***body**, **base**⟩ term is redundant as any interpretation on the **body** of a pseudopartitive in numeral free contexts could have been made on the **body** of the ⟨***body**, ***base**⟩ term from which it is derived.

[17] I am ambiguous as to whether O is the realisation of a classifier head, or whether it is a presupposition/case marking associated with a covert classifier in the Longobardi (1996) sense. Either way, I take the presence of O to indicate classificatory structure.

(66) Schematic of plural classifier semantics

$\textsc{cl}_{\textsc{plur}}(\mathrm{P}) = \langle {}^{*}\mathbf{body},\ {}^{*}\mathbf{base} \rangle \rightarrow \langle {}^{*}\mathbf{body},\ \mathbf{base} \rangle$

(67) Pseudopartitive as a realisation of $\textsc{cl}_{\textsc{plur}}(\mathrm{P})$:

MERCH	$\rightarrow$	MERCH-ED	$\rightarrow$	O FERCH-ED
girl	$\rightarrow$	girl-$\textsc{pl}_{\textsc{nc}}$	$\rightarrow$	girl-$\textsc{pl}_{\textsc{nc}}$-$\textsc{cl}_{\textsc{plur}}$
$\langle \mathbf{body},\ \mathbf{base} \rangle$	$\rightarrow$	$\langle {}^{*}\mathbf{body},\ {}^{*}\mathbf{base} \rangle$	$\rightarrow$	$\langle {}^{*}\mathbf{body},\ \mathbf{base} \rangle$

I now formalise the schematic in (66) as in (68). $\textsc{cl}_{\textsc{plur}}(\mathrm{P})$ applies to a number neutral uncountable i-set $\langle {}^{*}\mathbf{body},\ {}^{*}\mathbf{base} \rangle$ which has a **body** and **base** both closed under sum, extracts the **base**, then extracts the individuals of that **base**, and uses the resulting set to generate a new i-set with an overlapping **body** and a disjoint **base**. Structurally, $\textsc{cl}_{\textsc{plur}}(\mathrm{P})$ is taken to project over a $\langle {}^{*}\mathbf{body},\ {}^{*}\mathbf{base} \rangle$ i-set. The derivation in (69) and tree in Figure 6.7 exemplifies the case of a collective term as input for $\textsc{cl}_{\textsc{plur}}(\mathrm{P})$, and the derivation in (70) and tree in Figure 6.8 exemplify the case where a plural marked term serves as input for $\textsc{cl}_{\textsc{plur}}(\mathrm{P})$.

(68) $$\textsc{cl}_{\textsc{plur}}(\mathrm{P}) \rightarrow \lambda P. \begin{cases} \langle {}^{*}\mathrm{IP}\textsc{arts}_{\textsc{base}(\mathrm{P})},\ \mathrm{IP}\textsc{arts}_{\textsc{base}(\mathrm{P})} \rangle & \text{if P is } \langle {}^{*}\mathbf{body},\ {}^{*}\mathbf{base} \rangle \\ \perp & \text{otherwise} \end{cases}$$

(69) O HWYAID (duck.$\textsc{col}$-$\textsc{cl}_{\textsc{plur}}$ = ‘*ducks*’)

$\textsc{cl}_{\textsc{plur}}(\mathrm{P})(\textsc{hwyaid})$

$= \lambda P.\langle {}^{*}\mathrm{IP}\textsc{arts}_{\textsc{base}(\mathrm{P})},\ \mathrm{IP}\textsc{arts}_{\textsc{base}(\mathrm{P})} \rangle\ (\langle {}^{*}\textsc{hwyaid},\ {}^{*}\textsc{hwyaid} \rangle)$

$= \langle {}^{*}\mathrm{IP}\textsc{arts}_{(\textsc{base}(\langle {}^{*}\textsc{hwyaid},\ {}^{*}\textsc{hwyaid} \rangle))},\ \mathrm{IP}\textsc{arts}_{(\textsc{base}(\langle {}^{*}\textsc{hwyaid},\ {}^{*}\textsc{hwyaid} \rangle)))} \rangle$

$= \langle {}^{*}\mathrm{IP}\textsc{arts}_{({}^{*}\textsc{hwyaid})},\ \mathrm{IP}\textsc{arts}_{({}^{*}\textsc{hwyaid})} \rangle$

$= \langle {}^{*}\textsc{hwyaid},\ \textsc{hwyaid} \rangle$

☞ O HWYAID is a plural count i-set

It is plural, because $\textsc{body}_{(\textsc{o hwyaid})}$ is closed under *

It is countable, because $\textsc{base}_{(\textsc{o hwyaid})}$ is disjoint

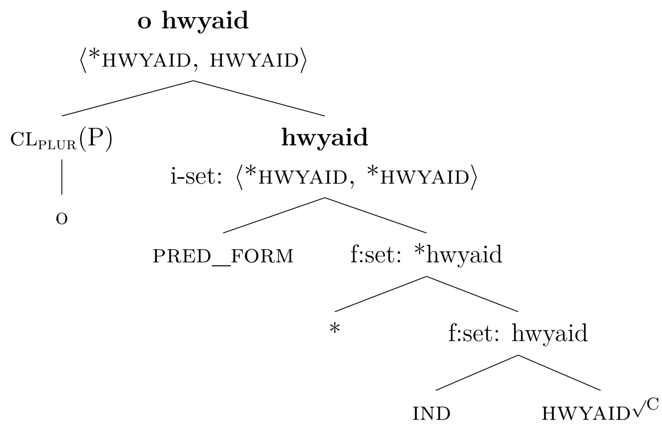


Figure 6.7: Welsh pseudopartitive i-set composition (collective input)

(70) O FERCH-ED (girl.SG-PL$_{NC}$-CL$_{PLUR}$ = '*girls*')

CL$_{PLUR}$(P)(MERCHED)

$= \lambda P.\langle{}^{*}\text{IPARTS}_{\text{BASE}(P)}, \text{IPARTS}_{\text{BASE}(P)}\rangle\ (\langle{}^{*}\text{MERCH}, {}^{*}\text{MERCH}\rangle)$

$= \langle{}^{*}\text{IPARTS}_{(\text{BASE}(\langle{}^{*}\text{MERCH}, {}^{*}\text{MERCH}\rangle))}, \text{IPARTS}_{(\text{BASE}(\langle{}^{*}\text{MERCH}, {}^{*}\text{MERCH}\rangle))}\rangle$

$= \langle{}^{*}\text{IPARTS}_{({}^{*}\text{MERCH})}, \text{IPARTS}_{({}^{*}\text{MERCH})}\rangle$

$= \langle{}^{*}\text{MERCH}, \text{MERCH}\rangle$

☞ O FERCHED is a plural non-count i-set

It is plural, because BODY$_{(\text{O FERCHED})}$ is closed under *

It is countable, because BASE$_{(\text{O FERCHED})}$ is disjoint

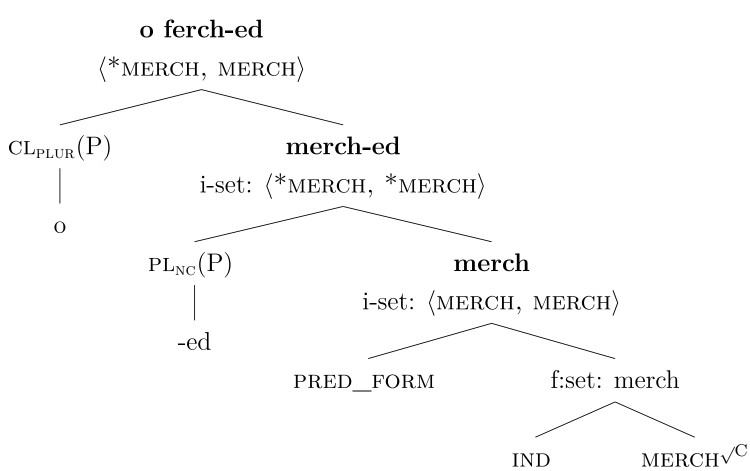


Figure 6.8: Welsh pseudopartitive i-set composition (plural input)

6.3.2 The (lack of a) plural of the singulative

Singulative marked terms in Welsh and Hejazi/Modern Standard Arabic are semantic singulars derived via $\text{CL}_{\text{SING}}(\text{P})$. The legal recursive stacking of functions predicts that singulative terms should be able to pluralise. This is because the singulative is a (derived) $\langle$**body**, **base**$\rangle$ term and so a well formed input for plural functions $\text{PL}_{\text{C}}(\text{P})$ and $\text{PL}_{\text{NC}}(\text{P})$. Indeed, if only $\langle$***body**, **base**$\rangle$ predicates can combine with numerals, then the singulative *must* pluralise, as its i-set type is incompatible with numerals.

In the case of Hejazi Arabic, this prediction is completely borne out. In Welsh, the singulative is subject to the covert last resort mechanism for combination with numerals, and the overt plural of the singulative is ruled out on economy grounds.

6.3.2.1 Arabic

The semantics for collectives, singulative functions, and plural functions in Hejazi Arabic predict the legal stacking of plural of the singulative structures as schematised in (71).

(71) Arabic recursive stacking for the plural of the singulative:

BAṬ	→	BAṬ-AH	→	BAṬ-A-AAT
duck	→	duck-CL_{SING}	→	duck-CL_{SING}-PL_{C}
$\langle$***body**, ***base**$\rangle$	→	$\langle$**body**, **base**$\rangle$	→	$\langle$***body**, **base**$\rangle$

Such a derivation is attested. Examples (72)-(74) demonstrate that this stacking in numeral contexts. Note, the phonetic realisation of the Arabic plural of the singulative differs across contexts. Every singulative term has a feminine sound plural of the singulative as in (72). Some, but not all, singulatives have broken plural forms. SHAJAR-AH (tree.COL-SING = *'tree'*) is a singulative which has a broken plural form, as in (73). There is no semantic difference between the use of the feminine sound plural of the singulative and a broken plural in numerical contexts. In contexts with numbers over ten, compulsory singular agreement occurs, as in (74) (see §2.3.1 for further details).

(72) Malik shaaf **khams-ah shajar-a-at** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 five-F tree-SING-PL in DEF-garden
'*Malik saw five trees in the garden*' = (73)

(73) Malik shaaf **khams-ah 'ashjaar** fi al-hadiqa — H. Arabic
see.PST.M.3 Malik five-F tree.SING.PL in DEF-garden
'*Malik saw five trees in the garden*' = (53a) = (72)

(74) Malik shaaf **khamstashr-ah shajar-ah-∅** — H. Arabic
see.PST.M.3 Malik fifteen-f tree-SING-PL
fi al-hadiqa
in DEF-garden
'*Malik saw fifteen trees in the garden*'

A full derivation for the Arabic plural of the singulative is given in (75), with hierarchical structure represented in Figure 6.9. The tree is labelled with phonological content for ease of exposition.

(75) 'ASHJAAR (tree.COL-ISING-PL = *'tree'*)
$= \text{CL}_{\text{SING}}(\text{P})(\text{SHAJAR})$
$= \lambda P.\langle \text{IPARTS}_{\text{BASE(P)}},\ \text{IPARTS}_{\text{BASE(P)}}\rangle\ (\langle {}^{*}\text{SHAJAR},\ {}^{*}\text{SHAJAR}\rangle)$
$= \langle \text{IPARTS}_{(\text{BASE}(\langle {}^{*}\text{SHAJAR},\ {}^{*}\text{SHAJAR}\rangle))},\ \text{IPARTS}_{(\text{BASE}(\langle {}^{*}\text{SHAJAR},\ {}^{*}\text{SHAJAR}\rangle))}\rangle$
$= \langle \text{IPARTS}_{({}^{*}\text{SHAJAR})},\ \text{IPARTS}_{({}^{*}\text{SHAJAR})}\rangle$
$= \langle \text{SHAJAR},\ \text{SHAJAR}\rangle$
Merge $\text{PL}_{\text{C}}(\text{P})$
$= \lambda \text{P}.\langle {}^{*}\text{BASE(P)},\ \text{BASE(P)}\rangle\ (\langle \text{SHAJAR},\ \text{SHAJAR}\rangle)$
$= \langle {}^{*}\text{BASE}_{\text{SHAJAR}},\ \text{BASE}_{\text{SHAJAR}}\rangle$
$= \langle {}^{*}\text{SHAJAR},\ \text{SHAJAR}\rangle$

☞ 'ASHJAAR is a plural count i-set
It is plural, because $\text{BODY}_{(\text{'ASHJAAR})}$ is closed under sum
It is countable, because $\text{BASE}_{(\text{'ASHJAAR})}$ is disjoint

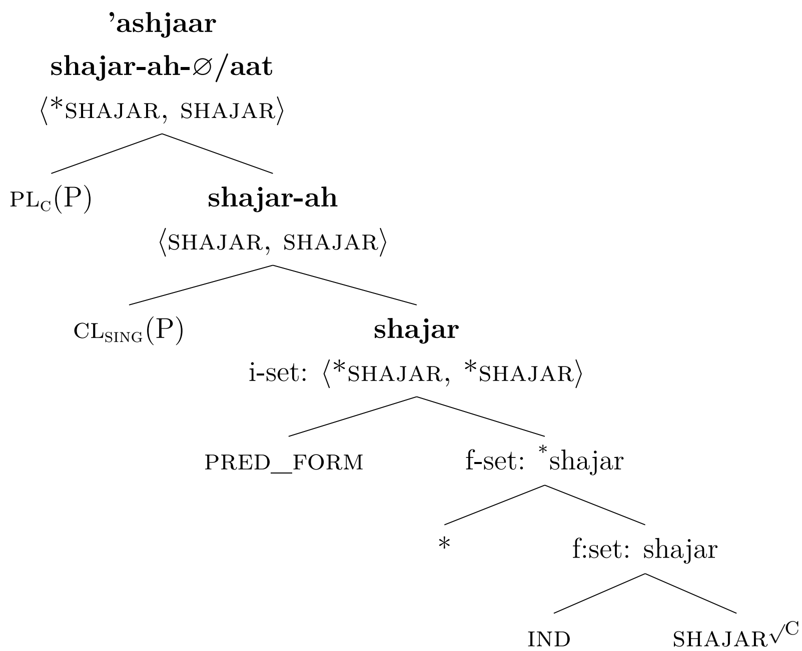


Figure 6.9: Arabic plural of the singulative i-set composition

The semantics given for CL$_{SING}$(P), PL$_C$(P) and recursive stacking can account for any form of the plural of the singulative in numerical contexts (before spellout considerations), but also the bare broken plural of the singulative; there is a reading of (76) as a standard exclusive plural (though see §2.3.1 for more details on interpretations of the broken plural of the singulative).

(76) Malik shaaf **'ashjaar** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 tree-SING.PL in DEF-garden
Plural reading: '*Malik saw trees in the garden*'
Abundance reading: '*Malik saw many trees in the garden*'

The semantics given for CL$_{SING}$(P), PL$_C$(P) and recursive stacking does not account for the paucal interpretation of the bare feminine sound plural of the singulative which is almost always interpreted as paucal, as in (9), repeated here for convenience.

(9) Malik shaaf **baṭ-a-at/shajar-a-at** H. Arabic
Malik see.PST.M.3 duck-SING-PAU/tree-SING-PAU
fi al-hadiqa
in DEF-garden
'*Malik saw a small number of ducks/trees in the garden*'

I follow Dali and Mathieu (2021b) in taking the (bare) Hejazi/Modern Standard Arabic feminine sound plural of the singulative to be a reflex of paucity. Therefore, the feminine sound plural of the singulative in bare contexts such as (9) is not a plural of the singulative in the sense that PL_{C}(P) has applied, but is in fact a the realisation of a distinct semantic function of paucity, which I call PAU(P).

(77) PAU(P) $\overset{\text{def}}{=}$ a semantic function that derives a paucal i-set

The ultimate claim for the Arabic system is that there are two plural-flavoured functions which may apply to semantic singulars: PL_{C}(P), and PAU(P). Both types of plural apply to semantic singulars, *i.e.* ⟨**body**, **base**⟩, whether they are subpredicatively derived (inherent singulars) or derived via function (individuating singulatives).[18] The canonical realisation of Arabic PL_{C}(P) is the broken plural form. It is therefore unsurprising that the broken plural of the singulative has no necessarily paucal nor exclusive semantics associated with it.[19] The canonical realisation of PAU(P) when applied to a singulative is the feminine sound plural (See §2.3.1 for more details on realisation of paucal and non paucal morphology).

[18]I do not offer a paucal semantics for the Subpredicative Iceberg Semantics model. Note though, the Subpredicative Iceberg Semantics model here may be amenable to a Harbour (2014); Martí (2017, 2020); Dali and Mathieu (2021b) inspired approach to paucity. In these approaches, paucal interpretations are built from plural terms, where complete join semilattices have have their upper bounds removed such that they are no longer complete. In theory, a paucal semantics for Subpredicative Iceberg Semantics may take the form of a function that applies to an established i-set with a **body** closed under sum in such a way that it remove its upper bounds. If one takes this view, then then the analysis for the paucal (of the singulative) would be such that it is built from the plural (of the singulative).

[19]In the case there is no broken plural form for a singulative term, then PL_{C}(P) cannot be realised when bare. This does not create any lexical gaps; the collective term covers meaning uses that bare PL_{C}(P) would have covered, the paucal covers uses when combining with numerals below ten, and PL_{C}(P) is covertly realised with numerals above ten.

6.3.2.2 Welsh

In theory, stacking of plural and singulative functions in Welsh should be legal: the singular count term should act as input for the plural function, PL_{NC}(P), deriving a plural of the singulative. Yet, Welsh has no plural of the singulative, in numeral contexts or bare: (11) and (78) are outright ungrammatical.

(11) *Gwelodd Owain **hwyad-enn-au** yn yr ardd — Welsh
see.PST.3 Owain duck-SING-PL in DEF garden
Intended: '*Owain saw ducks in the garden*'

(78) *Gwelodd Owain bum **hwyad-enn-au** yn yr ardd — Welsh
see.PST.3 Owain five duck-SING-PL in DEF garden
Intended: '*Owain saw five ducks in the garden*'

Under the Subpredicative Iceberg Semantics model being built here, the explanation is extremely simple. The Welsh plural function is PL_{NC}(P), which takes a singular count ⟨**body**, **base**⟩ i-set as input and then outputs a number neutral uncountable ⟨***body**, ***base**⟩ i-set. In the context of numerals as in (78), PL_{NC}(P) would never be grammatical due to the universal generalisation that ⟨***body**, ***base**⟩ i-sets cannot combine with numerals.

Though Welsh [numerals + PL_{NC}(P)] structures are ruled out generally, the overt plural of the singulative in Welsh never arises in the first place, even when bare. It is ruled out even before numerals are combined due to vacuity: PL_{NC}(P) derives a ⟨***body**, ***base**⟩ term and therefore applying it to a singulative term results in vacuous stacking, as schematised in (79). As part of more general Economy considerations, I invoke `NON-VACUITY`, as in (80).

(79) Welsh recursive stacking for the Welsh plural of the singulative (vacuous):

HWYAID	→	HWYAD-EN	→	HWYAD-ENN-AU
duck	→	duck-CL_{SING}	→	duck-CL_{SING}-PL_{NC}
⟨***body**, ***base**⟩	→	⟨**body**, **base**⟩	→	⟨***body**, ***base**⟩
				VACUOUS

(80) NON-VACUITY
Avoid functional structure that is semantically identical to a simpler form.

Ruling out the Welsh plural of the singulative via non-vacuity only applies to the overt plural of the singulative derived via $\text{PL}_{\text{NC}}(\text{P})$. Indeed, by hypothesis the covert Welsh $\text{PL}_{\text{C}}(\text{P})$ applies to singulatives terms in numeral environments as a non-vacuous last resort mechanism in much the same manner as the plural of the singular, but with a structure identical to the Arabic broken plural of the singulative in Figure 6.9.

Though the plural of the individuating singulative is ruled out, the plural of the packaging singulative is not, as in (81). Covert and overt plurals are correctly predicted to apply to singular count ⟨**body**, **base**⟩ i-sets derived via the packaging singulative, as there is no vacuity: the plural of a partition does not have the same mereological structure as a stuff mass term in either its **body** or **base**. Explicitly, the **base** of a stuff mass term is vertically overlapping, while the **base** of a partition is disjoint in context, and its plural is the closure under sum, which is not necessarily the same as the overlapping **base** of the stuff mass term. The plural of the packaging singulative is shown schematically in (82), with a full derivation in (83).

(81) Bwytodd Owain **gos-ynn-au** ddoe — Welsh
eat.PST.3 Owain cheese-PSING-PL yesterday
Owain ate pieces of cheese yesterday

(82) Schematic for the (overt) plural of the packaging singulative

COS	→	COS-YN	→	COS-YNN-AU
cheese.SMASS	→	cheese.SMASS-PSING	→	cheese.SMASS-PSING-PL_{NC}
⟨**body**, **base**⟩	→	⟨**body**, **base**⟩	→	⟨***body**, ***base**⟩
(overlapping **base**)		(disjoint **base**)		

(83) COS-YNN-AU (caws.SMASS-PSING-PL = *'pieces of cheese'*)
$= \text{PORT}(\text{P})(\text{CHEESE})$
$= \lambda \text{P}.\langle \pi_{\text{BASE(P)}},\ \pi_{\text{BASE(P)}} \rangle\ (\langle \text{CAWS},\ \text{CAWS} \rangle)$
$= \langle \pi_{\mathbf{base}(\langle \text{CAWS},\ \text{CAWS} \rangle)},\ \pi_{\mathbf{base}(\langle \text{CAWS},\ \text{CAWS} \rangle)} \rangle$
$= \langle \pi\text{CAWS},\ \pi\text{CAWS} \rangle$
Merge PL_{NC}(P)
$= \lambda \text{P}.\langle {}^*\text{BASE}(\text{P}),\ {}^*\text{BASE}(\text{P}) \rangle\ (\langle \pi\text{CAWS},\ \pi\text{CAWS} \rangle)$
$= \langle {}^*\text{BASE}_{\langle \pi\text{CAWS},\ \pi\text{CAWS} \rangle},\ {}^*\text{BASE}_{\langle \pi\text{CAWS},\ \pi\text{CAWS} \rangle} \rangle$
$= \langle {}^*\pi\text{CAWS},\ {}^*\pi\text{CAWS} \rangle$

☞ COSYNNAU is a number neutral uncountable count i-set
It is plural, because $\text{BODY}_{(\text{COSYNNAU})}$ is closed under sum
It is uncountable, because $\text{BASE}_{(\text{COSYNNAU})}$ is closed under sum

In summary, the semantics for the collective, individuating singulative, and overt plural in Welsh predict that stacking is functionally possible, but due to non-vacuity it is ruled out. Such constraints only hold for the individuating singulative, and so the plural of the packaging singulative is predicted, as is the use of the covert plural as a last resort mechanism in numeral contexts.

6.3.2.3 Excursus: Breton

Does non-vacuity apply only in Welsh, or is it a more broad constraint? If it is the case that non-vacuity is a crosslinguistic tendency, then the immediate consequence is that the plural of the singulative is ruled out in any singulative languages with uncountable plurals. This being the case, the elephant in the room is Breton.

First, consider that Breton, like its fellow Insular Celtic friends, has overt plurals (85), but uses the unmarked form is used in numeral constructions (86). The first conclusion is that Breton mirrors Welsh: the overt plural is a PL_{NC}(P) projection, and there is a covert PL_{C}(P) in numeral contexts.[20]

[20] The examples in this section are thanks to Steve Hewitt *p.c.*, but glossing is my own.

(84) Mael neus gweled ur **glesker** bah ar jardin — Breton
Mael be.M.PST see INDEF frog in DEF garden
'*Mael saw a frog in the garden*'

(85) Mael neus gweled **glesker-ed** bah ar jardin — Breton
Mael be.M.PST see frog-PL in DEF garden
'*Mael saw frogs in the garden*'

(86) Mael neus gweled **pemp glesker** bah ar jardin — Breton
Mael be.M.PST see five frog in DEF garden
'*Mael saw five frogs in the garden*'

Breton also has collective terms and a singulative marker, -ENN, as in (87) and (88), respectively. Like singular terms, when combining with numerals, the singulative form does not appear with an overt plural, as in (89).

(87) Mael neus gweled **logod** bah ar jardin — Breton
Mael be.M.PST see mouse in DEF garden
'*Mael saw mice in the garden* '

(88) Mael neus gweled ur **logod-enn** bah ar jardin — Breton
Mael be.M.PST see INDEF mouse-SING in DEF garden
'*Mael saw a mouse in the garden*'

(89) Mael neus gweled **pemp logod-enn** bah ar jardin — Breton
Mael be.M.PST see five mouse-SING in DEF garden
'*Mael saw five mice in the garden*'

The natural analysis for Breton again mirrors Welsh: the functional inventory contains an overt uncountable plural PL_{NC}(P) used in bare contexts, a covert countable plural PL_{C}(P) applies in numeral contexts as last resort, and there is a singulative classifier, CL_{SING}(P). If non-vacuity also holds in Breton, then the overt plural for the plural of the singulative should also be ruled out when bare. Yet, Breton is reported to pluralise its singulatives (Press, 2004; Hemon, 1975; Acuaviva, 2008). The available examples of the Breton plural of the singulative in the literature are few, and are not typically given in context. Two examples from Press (2004) are given in table Table 6.2.

Collective	**Singulative**	**Pl. of the singulative**	**Gloss**
stered	stered-enn	stered-enn-oú	star → a star → stars
deil	deli-enn	deli-enn-oú	foliage → leaf → leaves

Examples from Press (2004)

Table 6.2: The plural of the singulative in Breton

This seems to be direct evidence against non-vacuity. However, the status of the plural of the singulative in Breton is not clear cut. First, note that Press (2004) reports that the plural of the singulative in Breton is restricted, stating that it is:

> confined to particular words, and may be subject to dialectal variation. (Press, 2004, p.26)

Further, according to Steve Hewitt (*p.c.*) the plural of the singulative in Breton is not productive, is rarely attested, and has a weak paucal interpretation. The example in (90) is reported as not natural, and if accepted, it has a paucal interpretation.

(90) #Mael neus gweled **logod-enn-où** bah ar jardin — Breton
Mael be.M.PST see mouse-SING-PL in DEF garden
'*Mael saw a small handful of mice in the garden*'

It may well turn out that the [collective → singulative → plural of the singulative] three way contrast in Breton is overstated. There is some corpus evidence for this. For example, in Jean Lecoulant's ongoing corpus work on Breton[21], there is a STERED-ENN → STERED-ENN-OÚ contrast. At the time of writing, there is no evidence that the collective form STERED is present in the corpus. From discussion

[21] Jean Lecoulant's corpus is available here: https://kaojoubrezhoneg.wordpress.com

with Lecoulant (*p.c.*), it seems that for the speaker in the corpus, STEREDENNOÚ has replaced STERED entirely, and that finding a speaker with a productive three way contrast is rare. While this evidence is circumstantial, it may be the case that some nominal terms in Breton may be leaving the collective/singulative class as singulatives are reanalysed as singulars such that only a singular/plural contrast remains.[22]

The evidence from Breton is thin at this point, and lack of evidence for a productive plural of the singulative is not evidence of its absence. However, it stands to reason that the Breton plural of the singulative may very well be unproductive/crystallised; or perhaps not actually a plural of the singulative at all, but a plural of a reanalysed singular and the collective has fallen out of use; or that it has further semantic import (such as paucity). Any of these conclusions are compatible with non-vacuity: there seems to be no free application of the plural of the singulative in the same way that there is a plural of the singular.

***End of excursus**

6.3.2.4 Excursus: Mandarin Chinese

I take a small side-step from discussions of the (lack of) the plural of the singulative proper into a discussion of Mandarin Chinese, specifically the combining of numerals and classifiers, and also tentative discussion of reduplicated classifiers, which both show behaviours parallel to (ruled out) plural of the singulative.

First, recall that all nominal terms in Mandarin Chinese are $\langle$***body**, ***base**$\rangle$ i-sets, and that a classifier is required for the legal combination of numerals. Classifiers can come in general form, as GÈ in (91), or specific form, *e.g.* PIAN in (92). PIAN is typically associated with leaves and plantlife.

[22] There is perhaps a parallel here with the faux singulative class in Welsh which is a class of former and/or potential collective/singulative terms (See §2.1.5 for more details).

(91) Shīyáng zài huāyuán lǐ kàndào-le **wǔ gè nǚhái** M. Chinese
Shīyáng at garden in see.PST five CL girl
'*Shīyáng saw five girls in the garden*'

(92) He-li piao-zhe (yi) **pian shuye** M. Chinese
river-in float-DUR one CL leaf
'*There is a leaf floating on the river*'
NOT: '*There are many leaves floating on the river*'
From Zhang (2013, p.3)

Focussing on the general classifier, GÈ, I have already suggested a unification of its semantics with the singulative marker in Welsh and Hejazi/Modern Standard Arabic (§6.2.5). If this is the case, then there must be a covert PL_{C}(P) projection for the combination of classified terms with numerals, structurally parallel to the plural of the singulative (sans raising). I suggest that this PL_{C}(P) projection is not free, and is only used as a last resort in numerical constructions, and therefore is covert.[23]

What then, about overt plurals of classifiers? If the reasoning in §6.2.4 is on the right lines, then a covert last resort PL_{C}(P) arises because the overt plural is PL_{NC}(P) and cannot be used in numerical contexts. As it turns out, there are a few candidates for the overt PL_{NC}(P) in Mandarin Chinese, including the associative marker -MEN (Park, 2008; Cheng and Sybesma, 1999; Kim, 2008; Li, 1999; Kang, 1994; Iljic, 1994), plural classifier XIE (Iljic, 1994; Li, 1999; Zhang, 2013; Wu, 2019) and classifier reduplication (Zhang, 2013) (See footnote 48 in 2.3 for more details). Of these plurality inducing methods, the reduplication of classifiers is of most interest for the current purposes. When a classifier is reduplicated in Mandarin Chinese, a strictly exclusive plural reading arises where there is no reference to singularities, as in (93). The use of a numeral in these contexts is ungrammatical, as in (94). Note, the use of YI (*'one'*) in *e.g.* (93) is as an indefinite rather than a numeral, and its appearance is optional (see Zhang (2013) for details).

[23] I assume that the covert PL_{C}(P) may also apply to specific classifiers as last resort, but as I do not give a semantics for specific classifiers, I leave this as an assumption.

(93) He-li piao-zhe (yi) **pian-pian shuye**
river-in float-DUR INDEF CL-RED leaf
'*There are many leaves floating on the river*'
NOT: '*There is a leaf floating on the river*'
From Zhang (2013, p.3)

(94) *He-li piao-zhe **wu pian-pian shuye**
river-in float-DUR five CL-RED leaf
Intended: '*There are five leaves floating on the river*' [24]

The ungrammaticality of a [numeral + [reduplicated classifier]] structure indicates that reduplication of a classier has underlying uncountable plural semantics, *i.e.* it is a form of PL_{NC}(P) which derives a ⟨***body**, ***base**⟩ predicate. If this is the case, then non-vacuity should reasonably apply. Yet, this is not the case for (93), which is perfectly grammatical.

Notice, however, that PIAN is a specific classifier for terms of leaves and plantlife. In similar contexts, reduplication of the general classifier is completely ungrammatical when bare (95) or in the context of numerals (96). The only grammatical structure is one with the obligatory presence of the indefinite YI.

(95) *Shīyáng zài huayuán lǐ kandào-le **gè-gè nǚhái** M. Chinese
Shīyáng at garden in see-PST CL-RED girl
Intended: '*Shīyáng saw girls in the garden.*'

(96) *Shīyáng zài huayuán lǐ kandào-le wu **gè-gè nǚhái** M. Chinese
Shīyáng at garden in see-PST five CL-RED girl
Intended: '*Shīyáng saw five girls in the garden.*'

(97) Shīyáng zài huayuán lǐ kandào-le yi **gè-gè nǚhái** M. Chinese
Shīyáng at garden in see-PST INDEF CL-RED girl
'*Shīyáng saw girls in the garden.*'

[24] Example (94) is not in Zhang (2013), but is thanks to Shīyáng Fu.

These observational facts fall out from the functions of numbers proposed in this chapter in conjunction with non-vacuity. Consider that general classifiers offer no specific semantics other than individuation, while specific classifiers have associated meanings of size, shape, notional category (Chao, 1968; Doetjes, 2021; Nurmio, 2023). It stands to reason, then, that because the general classifier GÈ offers no semantics other than individuation, applying $\text{PL}_{\text{NC}}(\text{P})$ is necessarily vacuous, and is therefore ruled out by non-vacuity. During a [gè [gè [⟨***body**, ***base**⟩]]] derivation, the only semantic change is a matter of perspective where there is i-set reorganisation of (non-connected) sums of minimal individuals. Therefore, a bare [gè [gè [⟨**body**, **base**⟩]]] derivation is necessarily identical to the ⟨***body**, ***base**⟩ term from which it is derived, and is therefore ruled out.

For specific classifiers on the other hand, the picture is different. The use of a specific classifier, *e.g.* PIAN, is not just a simple re-organisation of (non-connected) sums, but imparts further semantics that are specifically related to the classifier. During a [pian [pian [⟨**body**, **base**⟩]]] derivation, not only is the perspective changed, but it is changed with respect to specific classifier semantics. Applying a $\text{PL}_{\text{NC}}(\text{P})$ will therefore result in an i-set that is not inherently identical to the ⟨***body**, ***base**⟩ source, and so non-vacuity does not apply. The picture for reduplicated classifiers is summarised in (98).

(98) Mandarin Chinese recursive stacking for reduplicated general classifiers:

SHUYE	→	PIAN SHUYE	→	PIAN-PIAN SHUYE
⟨***body**, ***base**⟩	→	⟨**body**, **base**⟩	→	⟨***body**, ***base**⟩
				Not vacuous
NǙHÁI	→	GÈ NǙHÁI	→	GÈ GÈ NǙHÁI
⟨***body**, ***base**⟩	→	⟨**body**, **base**⟩	→	⟨***body**, ***base**⟩
				Vacuous

An interesting point to note here is that a [gè [gè [⟨***body**, ***base**⟩]]] is not always ungrammatical: it may be saved by further semantic import. In the case of (97), the indefinite YI prevents the derivation from being identical to the initial ⟨***body**, ***base**⟩ i-set. This observation goes further. If the noun in a [gè [gè [⟨***body**, ***base**⟩]]] structure is topicalised as in (99), or if a distributive operator DŌU (*'all'*) is present, as in (100) and (101), then reduplication of GÈ is perfectly fine. It would therefore seem that only [gè [gè [⟨***body**, ***base**⟩]]] derivations where no further semantic import is induced are completely banned.[25]

(99) Lùluó-rén **gè-gè** fèn jī — M. Chinese
Lùluó-person CL-RED furious
'*People in Lùluó are all furious*'

(100) **gè-gè** Lùluó-rén dōu fèn jī — M. Chinese
CL-RED Lùluó-person all furious
'*As for all the Lùluó people, each of them is furious*'

(101) nà wǔ gè nǚhái Shīyáng zài huāyuán **gè-gè** dōu jiàn-le — M. Chinese
that five CL girl Shīyáng in garden CL-RED all see-PST
'*As the for 5 girls, Shīyáng saw each of them in the garden*'

***End of excursus**

6.4 Comments on inclusive/exclusive plurality

The system developed in this chapter is built on the assumption that plural semantics is inclusive, where interpretation of plural reference takes place in the **body** which is closed under Linkian *. Indeed, the inclusive approach to plurals is *necessary* to account for the lack of the plural of the singulative in Welsh. If plural markers did have an exclusive semantics, then non-vacuity could never apply, and an exclusive plural of the singulative in Welsh would be predicted, contrary to fact.

[25] The examples in (99)-(101) are thanks to Xiaozhou Zhang.

While this account derives the nuances of the plural of the singulative in Welsh and Arabic, it says nothing regarding how plural interpretations gain exclusive readings in upwards entailing contexts. Not only this, but there is nothing regarding why ⟨***body**, ***base**⟩ i-sets gain exclusive readings when derived via PL_{NC}(P) in Welsh (*i.e.* plural marked terms), but do not gain exclusive readings when they are derived subpredicatively (*i.e.* collectives).

While I do not offer an account for the calculation of exclusive plural interpretation here, I do offer some considerations of how exclusive readings are derived. In (scalar) implicature approaches to exclusive plural readings, the general consensus is that (enriched) singular semantics entail plural semantics in upwards entailing contexts, and this triggers implicature calculation such that exclusivity is derived. In downwards entailing environments, the entailment relations are reversed, and no implicature is derived (Ivlieva, 2013; Mayr, 2015; Sauerland et al., 2005; Spector, 2007; Zweig, 2009). The general process is outlined in (102) for exclusive plurals in upwards entailing contexts and (103) for inclusive plurals in downwards entailing (negative) contexts.

(102) Edward saw girls in the garden

Literal meaning: ‘*Edward saw one, or more than one, girl in the garden*’
Competitor: ‘*Edward saw exactly one girl in the garden*’
Calculation: The competitor entails the literal, so implicature calculation negates the competitor
Derived meaning: ‘*Edward saw one or more than on girl in the garden but it is not the case that Edward saw exactly one girl in the garden*’ = ‘*Edward saw two or more girls in the garden*’

(103) Edward did not see girls in the garden

Literal meaning: '*Edward did not see one, or more than one, girl in the garden*'

Competitor: '*Edward did not see exactly one girl in the garden*'

Calculation: The competitor does not entail the literal, so no implicature calculation is triggered.

No derived meaning: Literal meaning arises in context.

In, for example, Ivlieva (2013)'s model of inclusive/exclusive plural interpretations, there is a necessary implicature calculation with every instantiation of plural semantics. Incorporation of such a move is not clearly appropriate for the Subpredicative Iceberg Semantics model being built here: we must distinguish the types of plurals which are subject to implicature calculation. We should only want $\langle {}^*\mathbf{body}, {}^*\mathbf{base}\rangle$ and $\langle {}^*\mathbf{body}, \mathbf{base}\rangle$ i-sets derived via *optional* plural functions $\textsc{pl}_{\textsc{nc}}(P)$ and $\textsc{pl}_{\textsc{c}}(P)$ to be subject to implicature calculation; the $\langle {}^*\mathbf{body}, {}^*\mathbf{base}\rangle$ i-sets derived via subpredicative means (*i.e.* number neutral uncountable terms) should not undergo necessary implicature calculation. I suggest an amendment/clarification for incorporating compulsory implicatures into the Subpredicative Iceberg Semantics account. I suggest that a violation constraint such as (104) triggers implicature calculation. As such, only plurals derived directly from (enriched) singulars are subject to implicature calculation.

(104) Plural entailment restriction

Avoid structure where the projection of a head is entailed by its absence.

6.5 Discussion

The account put forward in this chapter has expanded the Subpredicative Iceberg Semantics framework to account for basic functions of number that apply in and across languages, as schematised in Table 6.3, which is updated from (62) as to include $\textsc{cl}_{\textsc{plur}}$(P). The functions of number are not restricted to a single syntactic position, and so recursive stacking is predicted.

Function	**Schematic**	**Attested**
$\textsc{cl}_{\textsc{sing}}$(P)	$\langle$***body**, ***base**$\rangle \rightarrow \langle$**body**, **base**$\rangle$	Welsh: singulative Arabic: singulative Mandarin Chinese: generalised classifier
$\textsc{cl}_{\textsc{plur}}$(P)	$\langle$***body**, ***base**$\rangle \rightarrow \langle$***body**, **base**$\rangle$	Welsh: pseudopartitive
$\textsc{pl}_{\textsc{c}}$(P)	$\langle$**body**, **base**$\rangle \rightarrow \langle$***body**, **base**$\rangle$	English: plural marker Arabic: plural marker Welsh: covert last resort mechanism Turkish: covert last resort mechanism Mandarin Chinese: covert last resort mechanism
$\textsc{pl}_{\textsc{nc}}$(P)	$\langle$**body**, **base**$\rangle \rightarrow \langle$***body**, ***base**$\rangle$	Welsh: overt plural marker Mandarin Chinese: reduplicated classier
Port(P)	$\langle$**body**, **base**$\rangle \rightarrow \langle$**body**, **base**$\rangle$	Welsh: packaging singulative

Table 6.3: Functions of number in Subpredicative Iceberg Semantics

The account posited here is similar to Dali and Mathieu (2021b); Grimm (2012a,b) and Mathieu (2012a, 2013, 2014) in that the treatment of the singulative is ultimately classifier-like, and is further similar to Mathieu's account specifically in the treatment of singulative markers and general classifiers as syntactically/semantically akin. However, the account proposed here is more powerful than these previous approaches to nominal number. The first major difference was discussed chapter 5, and concerned the compositional nature of number neutral uncountable terms.

The second major difference between the Subpredicative Iceberg Semantics account here and a more traditional approach to nominal number is specific to this chapter. That is, the Subpredicative Iceberg Semantics account alone correctly predicts that the plural of the singulative should be absent in Welsh, but not Hejazi/Modern Standard Arabic. In previous literature, this issue is not addressed, and is even incompatible with the frameworks. For example, Grimm (2012b)'s account outright predicts a Welsh plural of the singulative, contrary to fact (see §4.2.3.3), and Dali and Mathieu (2021b)'s account cannot clearly distinguish why the plural of the singulative is absent in Welsh, but not Arabic (see §4.3.2.2).

This prediction that there should be no Welsh plural of the singulative is borne out in Subpredicative Iceberg Semantics due to the uniquely flexible approach to plurality in the architecture of the system. This is because plurality is not taken to be a simple closure under (some variant of) Linkian sum, *, as in so-called 'mountain semantics' (see Landman (2020) for discussion). Rather, plurality comes in multiple forms, including at least countable and non-countable plural terms, which differ in the the disjointedness status of the **base**. With this flexibility, we allow for information in both the **body** and the **base** of any given i-set to be generally available throughout the computation of a plural term (subpredicative or via function), and therefore allow certain semantic functions to be sensitive to the status of the **body**, the **base**, or both.

It is precisely this flexible approach to plurality that allows for fine-grained differences between types of plurality. This separation of features of nominal number across the **body** and **base** allows Welsh, Arabic and Mandarin Chinese to have a unified semantics for singulativisation/classification, $\text{CL}_{\text{SING}}(\text{P})$, but for differences to arise due to the type of plurality employed at different compositional stages, but in a manor that is still minimal and predicative across languages. The system developed here uniquely allows for Subpredicative Iceberg Semantics to predict the presence of the (overt) plural of the singulative in Arabic due to the type of plural employed in Arabic, *i.e.* $\text{PL}_{\text{C}}(\text{P})$. The presence of the (overt) POTS, both bare and in numeral contexts, is taken to be a straightforward case of recursive stacking of number functions (though the facts are obscured by the paucal of the singulative). The absence of the overt plural of the singulative in Welsh is also predicted, due to the type of overt plural in Welsh, $\text{PL}_{\text{NC}}(\text{P})$ interacting with the non-vacuity constraint that disallows vacuous i-set derivation.

While this model of nominal number was motivated by the behaviours of the (absence of) the bare plural of the singulative in Welsh and Arabic, this flexible approach to number allows for prediction that languages which have uncountable plurals cannot apply that uncountable plural to singulative/classified terms and allow them to appear bare. This prediction so far seems to hold across languages that are not bone-fide collective/singulative languages, including the case of Mandarin Chinese reduplicated classifiers. The prediction is formalised under Subpredicative Iceberg Semantics assumptions as in (105).

(105) `PREDICTION ONE: BANNED DERIVATIONS`

$$\begin{array}{lllll} & \langle {}^*\mathbf{body}, {}^*\mathbf{base}\rangle & \to & \textsc{cl}_{\textsc{sing}}(\mathrm{P}) & \to & \textsc{pl}_{\textsc{nc}}(\mathrm{P}) \\ = & \langle {}^*\mathbf{body}, {}^*\mathbf{base}\rangle & \to & \langle \mathbf{body}, \mathbf{base}\rangle & \to & \langle {}^*\mathbf{body}, {}^*\mathbf{base}\rangle \end{array}$$

is banned under non-vacuity

A second prediction that falls out from this more flexible approach to nominal number, which is not clearly predicted by more traditional approaches, regards meaning/form correspondences in number constructions. From the available data seen in Turkish, Welsh, Mandarin Chinese, and marginally Breton, there seems to be testable prediction that if a language has an uncountable plural, then it will combine numerals with the singular-form of the noun, regardless of if the singular form is marked. This prediction is formalised under Subpredicative Iceberg Semantics assumptions in (106).

(106) `PREDICTION TWO:` $\texttt{PL}_{\texttt{C}}$`(P) AS LAST RESORT`

If a language has a $\textsc{pl}_{\textsc{nc}}(\mathrm{P})$ marker as its canonical plural marker, then a (covert) $\textsc{pl}_{\textsc{c}}(\mathrm{P})$ will apply as a last resort mechanism only in numerical contexts.

The prediction in (106) is unique to the Subpredicative Iceberg Semantics model. In a Mathieu (2012a, 2013, 2014) constructionist style approach, singulative marked terms in Welsh should not be countable, contrary to fact (see 4.3.2.1). Of course, one may posit an account more in line with the Dali and Mathieu (2021b) account and post allosemic rules that zero-mark the plural of the singulative in numeral constructions, but then the question remains: why is the bare plural of the singulative unattested in Welsh but not Arabic? Therefore, it would seem that while more traditional accounts may circumvent the marking issue in numerical constructions, they cannot do so in a way that easily predicts the lack of the Welsh plural of the singulative in bare contexts, while also ruling it in for Arabic. The Subpredicative Iceberg Semantics model here is unique in that the flexible approach to nominal number predicts both behaviours.

The predictions in (105) and (106) neatly account for the data in a way in which more traditional approaches to collective/singulative languages cannot, all which unifying collective/singulative systems and generalised classifier systems. This being said, it is not entirely clear how far this system can go. There is a tension in that there is a marginal plural of the singulative for Welsh artificial collections (§2.3.2), and the status of the Breton plural of the singulative is unclear at best. Still, the Subpredicative Iceberg Semantics model developed here offers testable predictions for the architecture of nominal number in such a way that allows greater flexibility over its predecessors.

6.6 Summary and next steps

This chapter has extended to Subpredicative Iceberg Semantics model by proposing functions of number regarding a unified singulative/classifier function, a portioning function, and two types of plurality. The semantics are worked out in such a way that allows for the singulative in Arabic to pluralise, but for the singulative in Welsh to never pluralise, unless as a covert last resort mechanism in numeral constructions. The account proposed here offers testable crosslinguistic predictions on the meaning-form correspondences of the functions of nominal number that goes beyond the collective/singulative class, but includes number marking languages (*e.g.* Turkish) but also generalised classifier languages (*e.g.* Mandarin Chinese).

Chapter 7

Building a system: Nominal flexibility

The Subpredicative Iceberg Semantics developed in the previous chapters has assumed a relationship between construed ontologies and linguistic ontologies such that perceptual construals inherently influence the structures in natural language. This has taken form in two universal claims. The first universal claim follows the lexicalist tradition of taking all liquid concepts to be ultimately mapped to stuff mass terms in natural language. The second universal claim is that terms which denote pre-theoretical individuals must have a projection of the IND filter in the subpredicative composition. Following this, there is crosslinguistic variation on further relationships between the construed ontologies and linguistic ontologies, where construals of local unmarkedness modulate the grammatical distinction between count and number neutral uncountable terms in `PRESUP`* languages (number marking and singulative languages), but `FREE`* languages have compulsory number neutral uncountable composition for individual denoting terms.

Conspicuously, the Subpredicative Iceberg Semantics account so far has discussed only initial/default nominal compositions (chapter 5) and higher functions of number (chapter 6). There has been no discussion of NOMINAL FLEXIBILITY, specifically no discussions of the so-called UNIVERSAL GRINDER phenomenon where prototypical count terms may be used in a stuff mass way, *e.g.* DOG in 'THERE IS DOG IN MY SOUP'. Any serious account of the count/mass distinction cannot ignore nominal flexibility, as the fundamental nature of any theory of nominal number must modulate between stuff and non-stuff reference of the same spatiotemporal individuals. The aim of this chapter is simple: I explore nominal flexibility as informed by collective/singulative languages, Welsh and Hejazi Arabic, with the view to incorporate a theory of nominal flexibility into Subpredicative Iceberg Semantics.

In §7.1, I introduce the universal grinder phenomenon and discuss its non-universality with respect to the absence of stuff-readings in Mandarin Chinese. I summarise two approaches to stuff-readings, the CONSTRUCTIONIST ROOT MEANING APPROACH (Borer, 2005a; Barner et al., 2009) and the LEXICALIST LAST RESORT APPROACH (Cheng et al., 2008; Rothstein, 2010, 2017).

In §7.2, I take a closer look at the behaviour of the Welsh and Hejazi Arabic collective/singulative systems in contexts which are amenable to stuff-readings. I conclude that terms for fruits and vegetables should have a special status in a theory of nominal flexibility, as they may always (where context is amenable) gain stuff-readings regardless of 'inherent' nominal category. This suggests a level of nominal flexibility in line with the constructionist root meaning approach. However, I also show that contra the constructionist root meaning approach, but in line with the lexicalist last resort approach, that collective terms for animals resist stuff-readings. Further, I show that stuff-readings are not always a last resort mechanism in Welsh.

In §7.3 I build a theory of nominal flexibility where multiple sources of nominal flexibility are proposed. The first source of nominal flexibility proposed is inspired by the root meaning approach: PERSPECTIVE SHIFT. I argue that perspective shift is a subpredicative process where certain concepts (such as fruit/vegetable) are systematically compatible with both individual and substance subpredicative composition. The second source of nominal flexibility is inspired by the last resort approach: a COUNT → MASS 'grinding operation', GRIND, which may apply freely in Welsh and English, context permitting. The final source of nominal flexibility is what I call ACCIDENTALLY HOMOPHONOUS ROOT CONCEPTS, is unique to the extension Subpredicative Iceberg Semantics laid out in this thesis. This is is where there are two distinct root concepts that are conceptually related and homophonous.

7.1 Setting the scene

7.1.1 The universal grinder

There is a degree of NOMINAL FLEXIBILITY afforded to nominal terms. Given appropriate morphosyntactic conditions and context, canonical count terms may be used in a stuff mass sense (COUNT → MASS nominal flexibility), and canonical stuff mass terms may be used in a count sense (MASS → COUNT nominal flexibility).

In the realm of MASS → COUNT nominal flexibility, there are two major archetypes of coercion that arise when prototypical mass terms are used in grammatically count contexts. THE UNIVERSAL SORTER is the phenomenon where a canonical stuff mass term is coerced into a reading of countable subkinds (Bunt, 1985). THE UNIVERSAL PACKAGER is the phenomenon where a canonical stuff mass term has a reading of countable contextually defined portions (Bach, 1986a; Jackendoff, 1991; Pelletier, 1975/1979). Examples of sorting and packaging are in (1). For the sorter reading, Edward bought three kinds of water (*e.g.* sparkling, mineral, flavoured) and for the packager reading, Edward bought multiple standard servings of water (*e.g.*. three bottles or three pints). Packaging is most frequently seen in restaurant contexts where local constraints restrict interpretation (Chierchia, 2010a; Landman, 2011; Rothstein, 2011, 2017, 2021). In this manner, context is crucial to packaging interpretation, as seen in (2) which is infelicitous unless uttered in the fanciful context of a vampire bar, where Edward, Damon and Angel are vampires and Edward bought a round of bloods for everybody.[1]

[1] I credit the conception of the vampire bar context to Chierchia (2010b, 2021).

(1) Edward bought three waters — English

Sorter: '*Edward bought three kinds of water*'
Packager: '*Edward bought three bottles/pints etc. of water*'

(2) # Edward bought three bloods — English

The discussion in this chapter focuses on THE UNIVERSAL GRINDER phenomenon, which Pelletier (1975/1979) credits to David Lewis.[2] The universal grinder is descriptively a COUNT → MASS coercion where canonical count terms are used in a stuff mass manner. For example, in normal plural use the count terms DOG and LION are interpreted as denoting multiplicities of integrated wholes: for (3a) multiple whole dogs are on the bed, and for (3b) multiple whole lions are in the zoo. These are canonical count readings where integrated wholes are referred to. Going forward I will use the term INTEGRAL-READINGS when discussing any interpretations that refers to pre-theoretical wholes. The term INTEGRAL-READING is intended to cover both singular count and number neutral uncountable readings: so long as integrated wholes are denoted, the reading is integral. Compare now DOG and LION in grammatical mass contexts as in (4). The interpretations are not of integrated wholes, but of dog matter/guts/sinew/innards *etc.* splattered all over the wall for (4a), and of lion meat in the soup for (4b). These are the readings that arise from the supposed universal grinder. Going forth, I will refer to the stuff mass interpretations of terms as in (4) as STUFF-READINGS.

(3) INTEGRAL-READING of DOG and LION — English

a. There are three **dogs** on the bed
b. There are three **lions** in the zoo

(4) STUFF-READING of DOG and LION — English

a. There is **dog** all over the wall
b. There is **lion** in my soup

[2] Grinding is rebaptized as FISSION in Landman (2011) and DOWNSHIFTING in Landman (2020).

An important aspect of stuff-readings in classical grinder literature is that grammatical change must be accompanied by physical change. This is argued for explicitly in Pelletier (1975)'s gruesome thought experiment, which I adapt here in context of (4a). Imagine placing a dog through a meat grinder, which chops and mashes the dog and spews out a homogenous mass of ex-dog stuff: meat, flesh, bones *etc.*. We then pick up that stuff and splatter it against a wall. After this event (4a) felicitous. Throwing a whole and healthy dog against a wall will not do if we wanted to secure the felicitousness of (4a). Rather, the poor dog must necessarily take a journey through the proverbial meat grinder. Similarly, (4b) is only compatible with an interpretation that there is lion meat in the soup, and not a whole lion. Again, physical change must have occurred for (4b) to be felicitous. The lion must be physically chopped up or ground down into it constituent parts, so to speak.[3]

With the grinding of dogs and lions, the conception of the universal grinder is undoubtedly grim, but the point is that the universal grinder is so-called as it is supposedly universal. Pelletier argues that, given the right context, any count noun can have stuff mass usage. In support of this, Pelletier cites Gleason (1969)'s famous termite example in (5), where BOOK and SHELF have stuff-readings as food, rather than integral-readings as integrated wholes.

(5) Termite family:
Johnny is very choosey about his food. English
He will eat **book**, but he won't touch **shelf**.
Example from Gleason (1965, p.136-137), emphasis my own.

[3]Note, stuff-readings are not the same as volume readings, which also arise in similar grammatical contexts. For example, in (i) there is a reading where no physical change is required to have happened to the apples (see Beviláqua et al. (2016); Landman (2011); Rothstein (2011, 2017) for more information.).

(i) Edward has more apple than Bella

Any serious discussion of the count/mass distinction must include a discussion of nominal flexibility, as the very existence of stuff-readings of canonical count terms requires a careful consideration of the mechanisms that underlie nominal structure and its interaction with construed ontologies. For `CONSTRUCTIONIST` approaches to nominal number, the source of integral- versus stuff-readings is purely syntactic, where stuff-readings are simply root meanings and integral-readings are derived via further syntactic structure (*e.g.* Bale and Barner, 2009; Borer, 2005a; de Belder, 2011). Alternatively, in the `LEXICALIST` tradition, nominal terms are either lexically count or lexically mass, and stuff-readings arise through a COUNT → MASS shifting device that is induced as a last resort mechanism in the case of a morphosyntactic mismatch (*e.g.* Cheng et al., 2008; Rothstein, 2010, 2017).[4]

7.1.2 The root meaning approach (constructionist)

The most well known constructionist account of stuff-readings is Borer (2005a)'s `ROOT MEANING APPROACH`. In this system, nominal terms begin life as acategorical roots which indiscriminately refer to 'stuff'. Count (integral) readings are derived via the projection of a divisional head, ***`div`***, and mass (stuff) readings arise when no ***`div`*** is projected. A schematic structure of a count term is shown in Figure 7.1. In this manner, Borer's approach follows Pelletier's wisdom: there is a stuff mass usage for every 'count' term, and hence the universal grinder phenomenon is truly universal; nothing stops a certain syntax composing, and the choice between mass and count structures is modulated by context.[5]

---

[4]See also Kagan (2024)'s contextual derivational suffix approach.

[5]A similar approach can be found in Pelletier (2012), who maintains the universality of the original Universal Grinder.

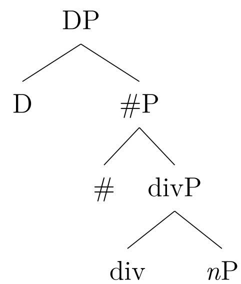


Figure 7.1: Borer's exoskeletal structure for nominal phrases

Borer's account of nominal number has been criticised for not being able to distinguish between number neutral uncountable and stuff mass reference, as both compositions are characterised by the lack of ***div*** (see §4.3 for more details). For example, Kagan (2024) criticizes constructionist approaches due to the fact that Slavic can induce stuff-readings via a suffix, which is not expected under any account where mass/stuff-readings arise via 'default' when no divisional head is projected.

An alternative constructionist account of stuff-readings as root meanings is found in Bale and Barner (2009). In Bale and Barner's system, there are multiple types of roots modelled including INDIVIDUATED SEMI-LATTICES which have (atomic) m-individuals as minimal parts and NON-INDIVIDUATED SEMI-LATTICES which have no minimal parts. In the syntax, there is a 'count' functional head which divides a non-individuated root into individuated singularities. Mass syntax is derived via identity function. Under this system, integral-reading are derived via the count functional head, and stuff-readings are derived via the identity function. For example, the root APPLE$^{\sqrt{c}}$ is taken to be a non-individuated root, and so the integral- versus stuff-readings arise through projection of either the individuating head or the identity function (essentially mass meanings are root meanings). Like Borer before them, Bale and Barner predict widespread nominal flexibility for integral- versus stuff-readings through syntactic means. To account for the difference between number neutral uncountable and stuff mass composition, Bale and Barner place the difference in the roots: FURNITURE is said to have a root structure that is individuated, and so the projection of the identity function head does not come out as 'stuff', but there is access to minimal parts.

7.1.3 The last resort approach (lexicalist)

The universality of the so-called universal grinder has been questioned. Stuff readings of canonical count terms have been accused of being jocular (see Gleason's termite example) and therefore limited. More worryingly for universalist accounts, stuff-readings are absent in some contexts, challenging that the universal grinder is universal in the sense it may apply to any 'count' noun (Beviláqua et al., 2016; Cheng et al., 2008; Chierchia, 2010b; de Belder, 2013; Doetjes, 2017, 2021; Rothstein, 2010, 2017, *a.o.*). An example in English of an absent stuff-reading is in (6), where there is no reading of triangle stuff being all over the wall, whatever that may mean.

(6) There is **triangle** all over the wall. English

Crosslinguistically, the picture is more dramatic, where the universal grinder is challenged as universal in the sense that it applies in all languages. Cheng et al. (2008) show that animal denoting terms in Mandarin Chinese such as GǑU (*'dog/s'*) are resistant to stuff-readings. In a perfectly grammatically stuff mass context such as (8), GǑU does not have a stuff-reading of dog meat being splattered against the wall as English DOG does in (4a), repeated for convenience. Rather, GǑU has only an integral-reading of a wallpaper pattern of pictures of numerous (whole) dogs on the wall. To get the stuff-reading, a compound is used, as in (9). From this, Cheng et al. conclude that the constructionist approach to nominal flexibility is too strong: the lack of of a classifier in (8) implies the lack of ***div***, and therefore a stuff mass reading should arise, and yet there is no stuff-reading reading for (8). Therefore the so-called universal grinder is said not to be universal in the sense that it does not apply across languages.

(4a) There is **dog** all over the wall English

(7) dì-shang dōu shi shuǐ — M. Chinese
floor-TOP all COP water
'*There is water all over the floor*'
From Cheng et al. (2008), adapted

(8) qiáng-shang dōu shi **gǒu** — M. Chinese
wall-TOP all COP dog
Integral: '*There are dogs all over the wall*'
*Stuff: '~~*There is dog all over the wall*~~'
From Cheng et al. (2008), adapted

(9) qiáng-shang dōu shi **gǒu-ròu** — M. Chinese
wall-TOP all COP dog-flesh/meat
'*There is dog-meat all over the wall*'
From Cheng et al. (2008), adapted

Against a constructionist view of nominal flexibility, Cheng et al. (2008) and later Rothstein (2010, 2017) argue that the difference between English and Mandarin Chinese in their availability to achieve stuff-readings for bare terms is due to English, but not Mandarin Chinese, invoking a last resort mechanism for stuff-readings of canonical count terms. In this `LAST RESORT APPROACH`, there is a type shifting COUNT → MASS lexical operation that applies only in cases where there is a morphosyntactic mismatch. As English has morphosyntactic number marking on count terms, the last-resort mechanism applies when a count term lacks its morphosyntactic number marking in relevant contexts, as in (4a). Now, as Mandarin Chinese has no morphosyntactic number marking and all nouns are grammatical when bare, the last-resort mechanism is not induced, and so Mandarin Chinese has no stuff-readings of canonically integral terms (*i.e.* number neutral uncountable terms). As such, (8) is correctly predicted to not have a stuff-reading.

The last resort approach to stuff-readings has crosslinguistic support. Consider Modern Hebrew. Example (10) shows that count terms appear with no morphosyntactic number marking. Based on this, Rothstein (2017) argues that stuff-readings of Modern Hebrew bare count nouns should never be triggered as they are always grammatical. The example in (11) shows that this is the case, where despite the context prejudicing stuff-readings, such readings are absent.[6]

(10) dani ra'a namer — Modern Hebrew
Dani saw.PST tiger
'*Dani saw a tiger*'
From Rothstein (2017, p.189)

(11) axarey ha-te'una haya **kelev** al ha-šulxan — Modern Hebrew
after DEF-accident was.M dog.M on DEF-wall
Integral: '*After the accident, there was a dog on the table*'
*Stuff: '~~*After the accident, there was dog on the table*~~'
From Rothstein (2017, p.189), adapted

While both Rothstein and Cheng et al. assume a last resort mechanism, it is not the case that they take all stuff-readings to be derived via last resort. For stuff-readings of fruit and vegetable terms, other mechanisms are invoked. Consider that in (12), Mandarin Chinese PÍNGGUǑ (*'apple'*) is interpreted as having a stuff-reading, referring to apple stuff, pulp, bits *etc.*. Cheng et al. suggest that stuff-readings can be invoked in Mandarin Chinese via `CONTEXT AND WORLD KNOWLEDGE`. In this way, the primary reading of (12) is the integral-reading of a wall-paper pattern of apples, but the stuff-reading is possible in the context where a child eats lots of apple and then vomits it all over a wall. Similarly, the food contexts of (13) -(16) highly prejudice and invoke the stuff-reading of various food-stuff nouns.

(12) Context: Child eats a lot of apple and then vomits on the wall:

qiáng-shang dōu shi **píngguǒ** — M. Chinese
wall-TOP all COP dog
Stuff: '*There is apple all over the wall*'
From Cheng et al. (2008, p.56)

[6] Rothstein (2017) notes that in colloquial Hebrew stuff-readings can arise in contexts with a morphological gender mismatch.

(13) shālā lǐ yǒu **júzi** M. Chinese
salad inside have orange
Stuff: '*There is orange in the salad*'
From Cheng et al. (2008, p.56), adapted

(14) Zhè-dào cài kěnéng baohán le **jiānguǒ** M. Chinese
This-CL dish may contain PST nuts
Integral: '*This product may contain nuts*'
Stuff: '*This product may contain nut stuff*'

(15) Tang-lǐ yǒu **luóbu** M. Chinese
soup-in have carrot
Integral: '*There's whole carrots in the soup*'
Stuff: '*There's carrot in the soup*'

(16) Bǎ qiehǎo de **jiānguǒ** chǎo wǔ fenzhong M. Chinese
BA chopped DE nuts fry five minute
Stuff: '*Fry chopped nuts for five minutes*'

The role of context and world knowledge is curious. In contexts that highly prejudice stuff-readings, animal terms in Mandarin Chinese are nonetheless resistant to stuff-readings. Rothstein reports that there is no stuff-reading for (17), but only an integral-reading where a whole animal is served. Cheng et al. report that (18) is impossible under an interpretation of pig meat being in the salad. However, *c.f.* (19) where Zhang (2013, p.20) reports the stuff-reading for YU (*'fish'*).

(17) women zuótiān chī le **yú** M. Chinese
we yesterday ate PST fish
Integral: '*We ate fish yesterday*'
From Rothstein (2017, 187), adapted

(18) #shālā lǐ yǒu **zhū** M. Chinese
salad inside have pig
*Stuff: '~~*There is pig meat in the salad*~~'
From Cheng et al. (2008, p.56)

(19) Wo **yu** bu chi le M. Chinese
1 fish NEG eat PST
Integral: '*I will not eat the whole fish anymore*'
Stuff: '*I will not eat the fish meat anymore*'
From Zhang (2013, p.20), adapted

That Mandarin Chinese animals term are resistant to stuff-readings when bare is quite robust, despite the exception in (19). The examples in (20)-(22) are the same food contexts as (14)-(16). These contexts are highly prejudiced for stuff-readings, and can induce such readings for terms for vegetation. And yet, in these contexts terms for commonly eaten animals, such as JI (*'chicken'*), must appear in a compound with ROU to get the stuff-reading. Without ROU, the sentences in (14)-(16) are ungrammatical.

(20) Zhè-dào cài kěnéng baohán le **jī-ròu** M. Chinese
This-CL dish may contain PST chicken-meat
Stuff: '*This product may contain chicken*'

(21) Tang-lǐ yǒu **jī-ròu** M. Chinese
soup-in have chicken-meat
Stuff: '*There's chicken in the soup*'

(22) Bǎ qiehǎo de **jī-ròu** chǎo wǔ fenzhong M. Chinese
BA chopped DE chicken-meat fry five minute
Stuff: '*Fry chopped chicken for five minutes*'

Cheng et al. (2008) and Rothstein (2010, 2017) differ on their accounts of why Mandarin Chinese animal denoting terms cannot gain stuff-readings but terms for vegetation can. For Rothstein (2017), the COUNT → MASS shifting device only applies in cases of grammatical mismatch. While Rothstein does not deny context as potentially relevant for stuff-readings, she suggests that terms for fruits and vegetables are AMBIGUOUS, and their denotation includes reference to both wholes and their constituent parts. Therefore, stuff-readings for fruit and vegetable terms are not COUNT → MASS shifts at all but a case of lexical ambiguity. Terms for animals are not lexically ambiguous, and so are constrained by the last resort mechanism. In this way, the universal grinder is restricted in so much as it certainly does not apply universally in any sense. Rather, the grinding COUNT → MASS grinding operation applies to singular count nouns only as last resort. Therefore, if a language lacks singular count nouns, then the COUNT → MASS grinder is absent in that language. Further, there are processes that look like grinding, but they are not cases of grinding at all (*e.g.* stuff-readings of fruits and vegetables).

Conversely, Cheng et al. (2008) suggest that context can, in theory, induce the COUNT → MASS shift for animal terms, but there is a lexical blocking effect where forms which may appear with a compound to get stuff-readings, such as ROU (*'meat'*), block forms induced by context. Therefore, animals terms are blocked from the grinder mechanism in *e.g.* (20)-(22), but as terms for vegetation have no compound, they are not blocked from the grinder mechanism in *e.g.* (12)-(16). In this way, there is a subtlety in Cheng et al. account: the COUNT → MASS operation is indeed present in Mandarin Chinese, but its use is restricted by context and blocking factors. Therefore, there is some argument that the COUNT → MASS is universal in the sense that all languages have it, and what is debated is the felicitousness of its use.

7.1.4 Singulative languages and grinding

The constructionist and lexicalist approaches to nominal flexibility take different stances on the universality of stuff-readings for any term in and across languages. Under the constructionist approach, 'the universal grinder' is the perfect moniker as nominal flexibility is truly predicted to be universal where every nominal term may be used in both an integral (count) and constituent stuff (stuff mass) way.

Conversely, the lexicalist approach does not predict universal nominal flexibility, restricting stuff-readings to predictable grammatical and contextual uses.

For both constructionist and lexicalist approaches, (some) stuff-readings turn out to be necessarily tied up in morphosyntactic number marking. If the root meaning approach is correct, then the lack of nominal number marking indicates a lack of `div` (or equivalent) structure, leading to stuff-readings. If the last resort approach is correct, we expect stuff-readings in cases of incongruent semantics/syntactic number marking. These baseline assumptions of each model make Welsh and Arabic excellent testing grounds for an investigation of the availability of stuff-readings due to the fact they both a.) lack an (overt) indefinite article, as shown in (23) and (24) and b.) have a minor number system, the collective/singulative system, where the singular term is morphologically marked with respect to a collective, as shown in Table 7.1.[7]

(23) Gwelodd Owain **ferch** yn yr ardd — Welsh
see.PST.3 Owain girl in DEF garden
'*Owain saw a girl in the garden*'

(24) Malik shaaf **bint** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 girl in DEF-garden
'*Malik saw a girl in the garden*'

	Singular → Plural	**Gloss**	**Collective → Singulative**	**Gloss**
Welsh	merch → merch-ed	girl → girls	hwyaid → hwyad-en	duck(s) → duck
Arabic	bint → bin-aat	girl → girls	baṭ → baṭ-ah	duck(s) → duck

Table 7.1: Welsh and Arabic singular/plural and collective/singulative systems

The lexicalist last resort approach and constructionist root meaning approach make different predictions for stuff-readings in singulative languages. Under the lexicalist last resort approach approach, both singular and singulative marked terms

[7] In this chapter, discussion relies on Hejazi Arabic rather than MSA.

should resist stuff-readings, even in contexts that prejudice such readings, as both singulative and singular terms are always grammatical when bare. That singulative marked terms are grammatical when bare is shown in (25) and (26). Under the constructionist approach singulars, but not singulatives, should never be able to gain stuff-readings. This is because, under a constructionist root meaning approach approach, singulars are unmarked, and so ambiguous between singular count and stuff mass compositions (projection or lack of *div*, respectively). Singulative marking, however, implies division (Mathieu, 2012a, 2013, 2014), and so the stuff mass interpretation should be impossible.

(25) Gwelodd Owain **hwyad-en** yn yr ardd — Welsh
see.PST.3 Owain duck-SING in DEF garden
'*Owain saw a duck in the garden*'

(26) Malik shaaf **baṭ-at** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 duck-SING in DEF-garden-GEN
'*Malik saw a duck in the garden*'

For collective terms, the outlook for both approaches again differs. The lexicalist last resort approach predicts that collectives should resist stuff-readings, as bare collectives are always grammatical when bare, as shown in (27) and (28). We may however expect stuff-readings for fruit and vegetable terms if an amenable context is invoked. Under constructionist root meaning approach, collective terms should uniformly have stuff-readings available, as the lack of singulative marking implies lack of division.

(27) Gwelodd Owain **hwyaid** yn yr ardd — Welsh
see.PST.3 Owain duck in DEF garden
'*Owain saw ducks/a duck in the garden*'

(28) Malik shaaf **baṭ** fi al-hadiqa — H. Arabic
Malik see.PST.M.3 duck in DEF-garden-GEN
'*Malik saw a duck/ducks in the garden*'

7.2 Stuff readings in Welsh and Hejazi Arabic

In this section I investigate the availability of stuff-readings for the singular/plural and collective/singulative systems of Welsh and Hejazi Arabic. The examples in this sections are purposely designed to induce stuff-readings. As such, the investigation relies heavily on food contexts. The reasoning for this is that if stuff-readings are absent even in contexts that highly prejudice them, then the lack of stuff-readings is argued to be grammatical, rather than contextual, in nature.

7.2.1 Singulative and singular terms

Singulative marked terms in Welsh and Hejazi Arabic are morphosyntactically derived singulars. Under the lexicalist last resort approach, as both Welsh and Hejazi Arabic lack an overt indefinite article, we may expect behaviour similar to that which Rothstein (2017) reports for Hebrew: stuff-readings never arise for bare singular and singulative terms, as there would never be a syntactic mismatch condition. However, we may expect terms for vegetation to gain stuff-readings in amenable contexts. Under the root meaning approach, singulative markers are classifiers in `div` (Mathieu, 2012a, 2013, 2014), and so they are necessarily individuated, ergo they should have integral-reading only. Contrastively, singulars have no clear morphosyntactic mode of division, and so in theory should be compatible with stuff-readings (zero marked terms may be ambiguous between projection of lack of division).

(29) Prediction of the lexicalist last resort approach:
Neither singular or singulative marked terms may gain grounds in Welsh or Hejazi Arabic

(30) Prediction of the constructionist root meaning approach:
Singular, but not singulative marked, terms may gain grounds in Welsh and Hejazi Arabic

In Hejazi Arabic, the last resort approach prediction is borne out: both singulative and singular terms do not gain stuff-readings. Singulars and singulative marked terms (that denote fleshy animals) are outright incompatible with stuff-readings. Even if the context is amenable, it speaks that stuff-readings cannot be forced. In the food contexts in (31), bare singular terms ASAD (*'lion'*) and MAA3IZ (*'goat'*) and bare singulatives BAṬ-AH (duck.COL-SING = *'duck'*) and BAQAR-AH (cow.COL-SING = *'cow'*) have only integral-reading of a whole (and probably small) animal sitting in the soup. There is no reading of the soup having meat/flesh in it. To get the stuff-reading of a singular term, a compound using LAHM (*'meat'*) must be used, as in (32). Note, singulative marked terms do not form a compound with LAHM (*'meat'*) to gain the stuff-reading, but rather the collective term is used.[8]

(31) a. fi **asad** fi shurbat-ii — H. Arabic
in lion in soup-my
Integral: '*There is a lion in my soup*'
*Stuff: '~~*There is lion meat/stuff in my soup*~~'

b. fi **maa3iz** fi shurbat-ii — H. Arabic
in goat in soup-my
Integral: '*There is a goat in my soup*'
*Stuff: '~~*There is goat meat/stuff in my soup*~~'

c. fi **baṭ-ah** fi shurbat-ii — H. Arabic
in duck-SING in soup-my
Integral: '*There is a duck in the soup*'
Stuff: '~~(There is duck meat/stuff in the soup)*~~'

d. fi **baqar-ah** fi shurbat-ii — H. Arabic
in cow-SING in soup-my
Integral: '*There is a cow in the soup*'
Stuff: '~~(There is cow meat/stuff in the soup)*~~'

[8] The data is thin here, as there are very few collective terms for fleshy animals in Arabic that are amenable to natural food contexts. In (32) .COL is added to the glosses of collective forms to aid the distinction from singular forms (both are morphologically unmarked).

(32) a. fi **lahm asad** fi shurbat-ii — H. Arabic
in meat lion in soup-my
'*There is lion meat in my soup*'

b. fi **lahm maa3iz** fi shurbat-ii — H. Arabic
in meat goat in soup-my
'*There is goat meat in my soup*'

c. fi **lahm baṭ** fi shurbat-ii — H. Arabic
in meat duck.COL in soup-my
'*There is duck meat in my soup*'

d. fi **lahm baqar** fi shurbat-ii — H. Arabic
in meat cow.COL in soup-my
'*There is goat meat in my soup*'

e. #fi **lahm baṭ-ah/baqar-ah** fi shurbat-ii — H. Arabic
in meat duck-SING/cow-SING in soup-my
Intended: '~~*There is duck/cow meat in the soup*~~'

Welsh does not experience the predictions of either the last resort approach or the root meaning approach. Both singular and singulative terms have stuff-readings available in amenable contexts. The singular count LLEW (*'lion'*) and singulatives ADER-YN (bird.COL-SING = *'bird'*) and MOCH-YN (pig.COL-SING = *'pig'*) in (33) have two interpretations. One interpretation is the stuff-reading: the soup contains lion/bird/pig-meat. The second interpretation is the integral-reading: there is a (probably small) whole lion/bird/pig in the soup. While both stuff-readings and integral-readings are possible, the stuff-reading is preferred in these contexts.

(33) a. Mae yna **lew** yn y cawl — Welsh
be.3 there lion in DEF soup
Stuff: '*There is lion meat in the soup*'
Integral: '*There is a lion in the soup*'

b. Mae yna **ader-yn** yn y cawl — Welsh
be.3 there bird-SING in DEF soup
Stuff: '*There is bird meat in the soup*'
Integral: '*There a bird in the soup*'

c. Mae yna **moch-yn** yn y cawl — Welsh
be.3 there pig-SING in DEF soup
Stuff: '*There is pig meat in the soup*'
Integral: '*There a pig in the soup*'

Note, as shown in (34), the use of Welsh singular and singulative marked terms for stuff-readings is not restricted to food contexts; contexts of violent material recomposition can also induce stuff-readings.

(34) Gwelais y ddamwain. Roedd hi'n ofnadwy. — Welsh
Nid oedd siawns gyda'r mochyn druan.
Roeddwn i'n sefyll wrth ochr y lon, ac yn sydyn roedd yna **waed**, **mochyn**, **coeden** a **char** ym mhobman.
'*I saw the crash. It was awful.*
The poor pig didn't stand a chance.
*I was standing at the side of the road, and then suddenly there was **blood**, **pig**, **tree**, and **car** everywhere.*'
Where: WAED (blood.SMASS = *'blood'*), MOCH-YN (pig.COL-SING = *'pig'*), COED-EN (tree.COL-SING = *'tree'*), CHAR (car.SG = *'car'*)

The Welsh singulative stuff-readings are unexpected under the root meaning approach, as there is division in the structure (the singulative marker). Both the Welsh singular and singulative stuff-readings are also unexpected under the last resort approach, as singulatives and singulars are perfectly grammatical in these contexts. Of course, we may well appeal to context under a last resort approach, but this begs the question: why should context induce stuff-readings in Welsh, but not Arabic?

7.2.2 Collective terms

As bare collective terms are always grammatical in Welsh and Hejazi Arabic, the lexicalist last resort approach predicts that collectives in their entirety are resistant to stuff-readings, except perhaps for collectives which denote vegetation (in amenable contexts). For the root meaning approach, collectives are characterised by the lack of division, and so the prediction is that all collectives may gain stuff-readings.

(35) Prediction of the lexicalist last resort approach:
Collective terms should never get stuff-readings (unless they denote vegetation)

(36) Prediction of the constructionist root meaning approach:
All collectives should gain grounds.

Collectives in Welsh align with the predictions of the last resort approach. Collectives which denote animals such as MOCH (pig.COL = *'pig(s)'*) and ADAR (bird.COL = *'bird(s)'*) do not allow stuff-readings, even in highly prejudiced food contexts. Example (37a)-(37b) have only integral-readings, where there are small birds/pigs sitting in the soup. To achieve the stuff-readings for Welsh fleshy animal collectives, alternative structures must be used. Welsh has two methods to achieve this. The first is a compound, as in (38a)-(38b). The second method is the shift to the singulative, as seen above in (33b)-(34). The requirement for compounding or singular shift for collectives to gain stuff-reading does not hold for collectives which refer to fruits and vegetables. These terms may easily achieve integral-reading or stuff-readings without using a compound or singulative form, as in (37d).

(37) a. Mae yna **adar** yn y cawl — Welsh
be.3 there bird in DEF soup
Integral: '*There are birds in the soup*'
*Stuff: '~~*There is bird in the soup*~~'

b. Mae yna **moch** yn y cawl — Welsh
be.3 there pig in DEF soup
Integral: '*There are pigs in the soup*'
*Stuff: '~~*There is pig in the soup*~~'

c. Mae yna **gnau** yn y cawl — Welsh
be.3 there nut in DEF soup
Integral: '*There are nut(s) in the soup*'
Stuff: '*There are chopped/blended nuts in the soup*'

d. Mae yna **foron** yn y cawl — Welsh
be.3 there carrot in DEF soup
Integral: '*There are carrots(s) in the soup*'
Stuff: '*There are chopped/blended carrots in the soup*'

(38) a. Mae yna **gig adar** yn y cawl — Welsh
be.3 there meat bird in DEF soup
'*There is bird meat in the soup*'

b. Mae yna **gig moch** yn y cawl — Welsh
be.3 there meat pig in DEF soup
'*There is pig meat in the soup*'

For completion note that non-fruit collectives also reject stuff-readings holds for insect collectives, too. The collective GWENYN (bee.COL = *'bees'*) in (39) only has a reading of multiple whole bees in the soup. However, I will not discuss insect collectives in detail as context may well a role here (bees are not typically used as ingredients in Welsh speaking cultures).

(39) Mae yna **wenyn** yn y cawl — Welsh
be.3 there bee in DEF soup
Integral: '*There are bee(s) in the soup*'
*Stuff: '~~*There are chopped/blended bees in the soup*~~'

The behaviour of bare collectives in Hejazi Arabic are a different story, where neither the last resort approach nor the root meaning approach is clearly supported. Some collective terms for animals may gain stuff-reading readings, others may not. For example, BAQAR (cow.COL = *'cow(s)'*) in (40a) resists the stuff-reading. If (40a) is accepted at all, the integral-reading arises: there are multiple little cows sitting in the soup. To achieve the stuff-readings of BAQAR, a compound using LAHM (*'meat'*) is used, as shown in (32d), repeated. Note, unlike Welsh there is no shift to the singulative to induce stuff-readings. Similar observations are reported for BAṬ (duck.COL = *'duck(s)'*). Though this is the case for BAQAR and BAṬ, different observations are seen for DEJAAJ (chicken.COL = *'chicken(s)'*) and SAMAK (fish.COL = *'fish(es)'*), which allow stuff-readings when bare. The use of LAHM with SAMAK and DEJAAJ to form a compound is infelicitous. Finally, collectives that denote fruits and vegetables do not appear in compounds and allow stuff-reading when bare, as in (40e)-(40f).

(40) a. fi **baqar** fi shurbat-ii — H. Arabic
in cow in soup-my
Integral: '*There are cow(s) in the soup*'
*Stuff: '~~*There is cow meat in the soup*~~'

b. fi **baṭ** fi shurbat-ii — H. Arabic
in duck-SING in soup-my
Integral: '*There are duck(s) in my soup*'
Stuff: '~~(There duck meat in the soup)*~~'

c. fi **dejaaj** fi shurbat-ii — H. Arabic
in chicken in soup-my
Integral: '*There are chicken/s in the soup*'
Stuff: '*There is chicken meat in the soup*'

d. fi **samak** fi shurbat-ii — H. Arabic
in fish in soup-my
Integral: '*There are fish/es in the soup*'
Stuff: '*There is fish meat in the soup*'

e. fi **jazar** fi shurbat-ii H. Arabic
in carrot in soup-my
Integral: '*There are carrots(s) in the soup*'
Stuff: '*There are chopped/blended carrots in the soup*'

f. fi **basal** fi shurbat-ii H. Arabic
in onion in soup-my
Integral: '*There are onion(s) in the soup*'
Stuff: '*There are chopped/blended onions in the soup*'

(32) a. fi **lahm baṭ** fi shurbat-ii H. Arabic
in meat duck in soup-my
'*There is duck meat in my soup*'

b. fi **lahm baqar** fi shurbat-ii H. Arabic
in meat cow in soup-my
'*There is cow meat in my soup*'

(41) a. #fi **lahm dejaaj** fi shurbat-ii H. Arabic
in meat chicken in soup-my
Intended: '~~*There is chicken meat in my soup*~~'

b. #fi **lahm samak** fi shurbat-ii H. Arabic
in meat fish in soup-my
Intended: '~~*There is fish meat in my soup*~~'

For completion, like in Welsh, Arabic insect collectives also reject stuff-readings. The collective NAHL (bee.COL = *'bees'*) in (42) only has a reading of multiple whole bees in the soup. Again, I do not discuss this in detail, as context may well a role here too.

(42) fi **nahl** fi shurbat-ii H. Arabic
in bee in soup-my
integral-reading '*There are bees(s) in the soup*'
*Stuff: '~~*There are chopped/blended bees in the soup*~~'

Summary of observations

Welsh and Hejazi Arabic are not uniform in the strategies used to gain stuff-readings, nor are their behaviours clearly compatible with either the root meaning approach or the last resort approach. There is only one clear similarity: collective terms for fruits and vegetables, like general number terms in Mandarin Chinese, seem to be special and may easily gain stuff-readings. This apparent specialness of terms for vegetation holds even in English: uncountable fruit and vegetable terms may be substituted in the examples in (43), leading to ambiguity between integral-readings and stuff-readings, but uncountable terms for non-vegetation cannot be legally substituted for stuff-readings.

(43) a. There's ______ in my smoothie — English
b. There's ______ in the salad
c. This product may contain ______
Legal substitutions for integral-readings and stuff-readings:
FRUIT, CELERY, SPINACH, ROCKET, KALE, ASPARAGUS, OKRA
Illegal substitutions for stuff-readings:
*POULTRY, *CATTLE

It seems that there is systematic constituent stuff/integral ambiguity for terms which denote fruit and vegetables.[9] This is not a controversial claim. Landman (2011), for example, assumes that there is a count/mass ambiguity for fruit/vegetable terms in Mandarin Chinese, English and Dutch. Similarly, Rothstein (2017, p.187) holds that fruits and vegetable terms are unique in that they allow reference to both individuals and stuff. Finally, experimental evidence from Beviláqua et al. (2016) suggests that the stuff-reading of food terms is more readily available than for non-food terms in volume constructions.

[9]Recent discussion on Russian by Kagan (2024) seems to support this idea, though the idea of systematic ambiguity is not discussed. Kagan, p.43 points out that Russian collective GOROX (pea.COL = *'peas'*) also has reference to pea parts as well as pea wholes.

On the other hand, there is no crosslinguistic systematic stuff/integral ambiguity for terms which denote fleshy animals, even if the context is amenable to such readings. Appealing to grammatical category and morphosyntactic mismatches to explain the patterns is difficult. Singular count and singulative terms may gain stuff-readings in English and Welsh, but the data so far indicates that English stuff-readings are triggered by grammatical mismatches, but in Welsh they are not (I return to this point later in §7.3.2). For number neutral uncountable terms (collectives, general number), there seems to be a tendency for non fruit/vegetable terms to resist stuff-readings. Yet, to suggest a ban on grinding number neutral uncountable terms seems unintuitive, as stuff-readings are available for some Arabic collectives.

It is clear, then, that the so-called universal grinder phenomenon is multifaceted, with intertwining considerations that include, at the very least, notional classes (fruits, vegetables), but also grammatical classes (why should a non-fruit/vegetable collectives require singulativisation in *e.g.* Welsh to gain stuff-readings?) The observations that I will try to account for in the following sections are summarised in (44).

(44) `OBSERVATIONS`

a. There is systematic crosslinguistic stuff/integral ambiguity for terms which denote fruits and vegetables.

b. It is not necessarily the case that stuff-readings must be derived via last resort (in Welsh).

7.3 A subpredicative approach to stuff-readings

In what follows, I will offer an account of nominal flexibility both in and across languages. The account builds upon the foundations of Subpredicative Iceberg Semantics laid out in chapter 5 and 6. The account I will propose adopts a multiple source model of nominal flexibility, taking inspiration from both the constructionist root meaning approach and lexicalist last resort approach to stuff-readings. The account put forward is a hybrid approach to stuff-readings, where there are both structural mechanisms and COUNT → MASS devices, each responsible for different types of nominal flexibility.

The first source of nominal flexibility is `PERSPECTIVE SHIFT`. Taking inspiration from constructionist accounts (*e.g.* Bale and Barner (2009)), integral-readings and stuff-readings are derived via different configurations of subpredicative structure. I argue that perspective shift underlies the integral-readings and stuff-readings of 'inherently flexible' nouns like STONE, but also fruit and vegetable terms. Perspective shift can be seen as a direct interface of how construed ontologies influence linguistic ontologies at the subpredicative level.

The second source of nominal flexibility is more in line with the lexicalist tradition. A COUNT → MASS operation, which I call GRIND, is put forward. Following Rothstein (2017), I do not deny that irreconcilable morphosyntax may trigger GRIND. However, I add the caveat that some languages (such as Welsh) may freely trigger GRIND, though context determines its felicitousness. GRIND may be seen as a non-universal language specific shifting device where grammatical change necessarily reflects physical change.

The third source of 'nominal flexibility' is not nominal flexibility at all, but cases of `ACCIDENTALLY HOMOPHONOUS ROOT CONCEPTS`. For this type of 'flexibility', there are homophonous root concepts that are distinct at the linguistic ontological level, though their referents are structurally related at the construed ontological level.

In this manner, I ultimately argue there is no 'universal grinder' in and across languages. That is, there is no one single operation that can freely apply to all nouns in a given language. Rather, stuff-readings are derived across languages in a multitude of ways. Further, I will eventually claim that perspective shift is the only universal mechanism across all human languages.[10]

[10] Compounding for stuff-readings is assumed universal, but I do not discuss in detail here.

7.3.1 Perspective shift

The first source of stuff-readings/integral-reading ambiguity is `PERSPECTIVE SHIFT`, which occurs where a nominal term may refer to a given individual as an integrated whole or to its material constitution naïvely conceptualised as homogenous stuff of the same property.[11] A good example of perspective shift is shown by the senses of English STONE. When interpreted as a stuff mass term, as in (45), the stuff-reading arises. In this context, the use of a plural marker and numerals are infelicitous. Alternatively, STONE may be used as a count term, as in (46), where the integral-reading arises. In this context, plural markers and numerals are felicitous.

(45) There is **stone** in the road — English

  a. *There is stones in the road
  b. *There is three stones in the road

(46) There are **stones** in the road — English

  a. There are three stones in the road

On an intuitive level, the stuff-readings for STONE and DOG are different; while the dog must have necessarily undergone physical change in (4a), repeated, no physical change is needed for STONE. The interpretations of (45) and (46) may apply to the same configuration of stone-individuals. All that is changed is the *perspective* upon which we describe the individuals.

(4a) There is **dog** all over the wall

Accounting for perspective shift in the Subpredicative Iceberg Semantics model is simple in conception; de Vries and Tsoulas (2021) suggest that perspective shift is subpredicative, occurring at the level of the interaction between root concept and

[11] Perspective shift intends to capture the behaviour of `DUAL LIFE NOUNS` in the Kiss et al. (2021) sense, and `AMBIGUOUS NOUNS` in the Chierchia (2010b, 2017, 2021) sense.

filter. The basic idea is as follows. In theory, the subpredicative SUB and IND filters can technically apply to any root concept, but in some cases the filters will come out empty. There are nominal roots which are compatible with the SUB filter only (which I call $\sqrt{\text{SUB}}$ roots), which will come out as empty when combining with the IND filter. An example of a $\sqrt{\text{SUB}}$ root is WATER$^{\sqrt{c}}$. There are roots compatible with the IND filter only (which I will call $\sqrt{\text{IND}}$ roots), and so combining them with the SUB filter they would come out empty. An example of an $\sqrt{\text{IND}}$ root is DOG$^{\sqrt{c}}$. Perspective shift occurs where a given root concept is compatible with either the IND filter or the SUB filter (which I will call$\sqrt{\text{SUB/IND}}$ roots). STONE$^{\sqrt{c}}$ is a $\sqrt{\text{SUB/IND}}$ root, as is compatible with both subpredicative filters. The integral-reading reading of STONE arises through the use of the IND filter as in the tree in Figure 7.2, and the stuff-readings arises through the stuff mass composition (using the SUB filter) as in the tree in Figure 7.3.

(47) a. $\sqrt{\text{SUB}}$ root: compatible only with the SUB filter
roots for pre-theoretical liquids *e.g.* WATER$^{\sqrt{c}}$

b. $\sqrt{\text{IND}}$ root: compatible only with the IND filter
roots for pre-theoretical individuals *e.g.* DOG$^{\sqrt{c}}$

c. $\sqrt{\text{SUB/IND}}$ root: compatible with both the SUB filter and the IND filter
roots for pre-theoretical individuals with naïvely homogenous material composition *e.g.* STONE$^{\sqrt{c}}$

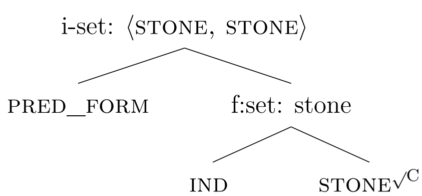


Figure 7.2: English STONE (singular count) i-set composition

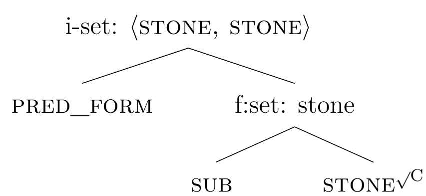


Figure 7.3: English STONE (stuff mass) i-set composition

The claim I put forward is a subpredicative version of Rothstein (2017)'s wisdom: fruit and vegetable stuff-readings are due to fruit and vegetable root concepts containing information regarding both the integrated whole and the constituent matter. In the Subpredicative Iceberg Semantics fashion, fruit and vegetable terms are composed from $\sqrt{\text{SUB/IND}}$ concepts, leading to a stuff mass and integral ambiguity. In the case of *e.g.* English, CARROT is systematically ambiguous between singular count and stuff mass composition, paralleling the case for STONE in the trees in Figures 7.2 and 7.3. In collective/singulative languages (and also generalised classifier languages), the minimal difference is that the ambiguity is between stuff mass and number neutral uncountable composition. This is shown for Welsh MORON as in the trees in Figures 7.5 and 7.4.

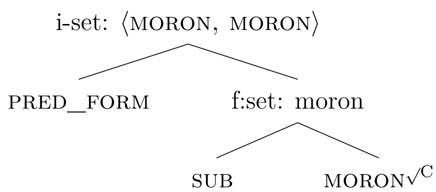


Figure 7.4: Welsh MORON (stuff mass) i-set composition

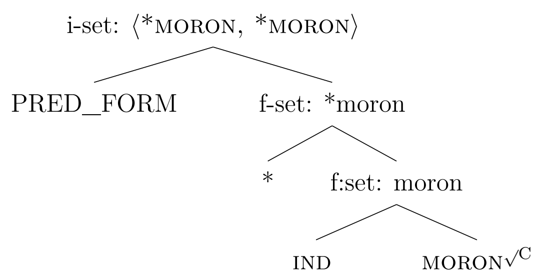


Figure 7.5: Welsh MORON (collective) i-set composition

7.3.1.1 Subpredicative filters

I now discuss the mechanisms of Subpredicative Iceberg Semantics such that perspective shift applies for percepts such as CARROT but not DOG. To start, let us look closer at the subpredicative filters. We start first with the IND filter, which de Vries and Tsoulas (2021) define as in (48). According to this definition, the IND filter combines with a root concept and outputs a filtered set of the pre-theoretical individuals relevant to that root concept.

(48) THE IND FILTER
IND $\stackrel{\text{def}}{=} \lambda C \lambda x[C(x) \wedge$ INDIVIDUAL(x)]
From de Vries and Tsoulas (2021, p.13)

What does it mean to be an individual? How can we ensure that the IND filter combines with CAT$^{\sqrt{c}}$ but not MILK$^{\sqrt{c}}$? A first approximation may be to incorporate the mereotopological notion of being maximally strongly self connected (MSSC) with respect to a property.[12]

(49) MAXIMALLY STRONGLY SELF CONNECTED WITH RESPECT TO A PROPERTY
MSSC(x,P) $\stackrel{\text{def}}{=}$ P(x) $\wedge$ SSC(x) $\wedge$ $\forall$y[P(y) $\wedge$ SSC(y) $\wedge$ $\mathcal{O}$(y,x) $\rightarrow$ y $\sqsubseteq$ x]
An m-individual is maximally strongly self connected relative to a property if (i) every part of the m-individual is connected (overlaps) the whole (strongly self connected) and (ii) anything which has the same property, is strongly self-connected, and overlaps it is once again part of it (maximality).[13]

This is not enough. Both a duck and a bikini are intuitive wholes that count as individuals. However, while any given duck is an MSSC whole, a given bikini is not an MSSC m-individual, but a scattered whole. The semantics of the IND filter must therefore be sensitive to different types of wholes. This involves at a minimum MSSC and scattered wholes, but may also include other types of wholes such as groups (*e.g.* COMMITTEE). To capture this, I model the property of being an individual as a series of mutually exclusive disjunctions and in (50), where MSSC(x,P) is a maximally strongly self connected m-individual with respect to property P, SCATTER(x,P) is a scattered whole with respect to a property P, and so on for different types of wholes.[14] With this, if the IND filter combines with DUCK$^{\sqrt{c}}$, then the output will be a filtered set of entities that satisfy the root concept DUCK$^{\sqrt{c}}$ that are also individuals, in this case, MSSC m-individuals.

[12]For an overview of the mereotopological notions I assume see §section 3.4.
[13]The translations for (49) is taken directly from Wągiel (2021b, p.198)
[14]I do not discuss the semantics of scattered wholes and groups here, but present for a more holistic picture.

(50) `INDIVIDUAL`

$\textsc{individual}(x) \stackrel{\text{def}}{=} \forall P[P(x) \rightarrow (\textsc{mssc}(x,P) \vee \textsc{scatter}(x,P) \ldots)]$

x is an individual if. for any property P x is maximally strongly self connected relative to property P, or is a scattered whole relative to property P ...

Turning now to the SUB filter, which de Vries and Tsoulas define as in (51).

(51) `THE SUB FILTER`

$\textsc{sub} \stackrel{\text{def}}{=} \lambda C \lambda x[C(x) \wedge \textsc{substance}(x)]$

From de Vries and Tsoulas (2021, p.13)

To define SUBSTANCE(x), I follow Grimm (2012b)'s intuition that any given instance of a substance implies substantial overlap with another m-individual of that substance. To represent this, like Grimm, I use Casati and Varzi (1999); Varzi (2007)'s mereotopological notion of `FIRM CONNECTEDNESS` (FC) in (53) as a key ingredient to define what it means to be a substance. With this, if the IND filter combines with $\textsc{water}^{\sqrt{c}}$, then the output will be the filtered set of entities that satisfy $\textsc{water}^{\sqrt{c}}$ that are also substances, in this case water m-individuals which are firmly connected to other water m-individuals. Note, the filtered set will necessarily be overlapping. Assuming x is a water m-individual with proper parts $a \sqcup b \sqcup c$ that are all water m-individuals, and x is in $\textsc{water}^{\sqrt{c}}$, then the filtered set will output all the water parts that are firmly connected to other water parts, including $a \sqcup b$ and $b \sqcup c$, which are vertically overlapping.

(52) `FIRMLY CONNECTED`

$\textsc{fc}(x,y) \stackrel{\text{def}}{=} \exists w \exists z[w \sqsubseteq x \wedge z \sqsubseteq y \wedge \textsc{ssc}(w \sqcup z)]$

Two m-individuals x and y are firmly connected if a sum of their parts is strongly self connected.

From Wągiel (2021b, p.200), adapted, which is adapted from Varzi (2007, p.62)

(53) SUBSTANCE
SUBSTANCE(x) $\stackrel{\text{def}}{=} \forall$P[P(x) $\wedge$ $\exists$y[P(y) $\wedge$ x $\neq$ y $\wedge$ FC(x,y)]
x is a substance if. for any property P x is firmly connected to a distinct m-individual of the same property.

As mereotopology is blind to the structure of matter, the SUB filter will easily compose with root concepts not only for liquid substances but also materials of matter. For example, GOLD$^{\sqrt{c}}$ will combine with the SUB filter to output all firmly connected instances of gold. The SUB filter will not output pre-theoretical individuals. This is because *e.g.* whole ducks are not firmly connected to other whole-ducks. Nor will the SUB filter output firmly connected parts of intuitive wholes. For example, *klaus'-tail*$\sqcup$*klaus'-body* are firmly connected, as is *the-top-half-of-the-water-in-my-glass*$\sqcup$*the-bottom-half-of-the-water-in-my-glass*, but only the water m-individuals count as SUBSTANCE(x), as they are firmly connected and of the same property. Therefore, if the SUB filter does combine with DUCK$^{\sqrt{c}}$ or CAT *etc.*, it will come out empty.

7.3.1.2 Root concepts and ontological mappings

Mereotopology is blind to the substance of matter. The apple on my table is as a good an MSSC m-individual as the water in my glass. The apple on my table counts as an MSSC m-individual so long as anything that is connected to it is part of the apple-whole. So too does the water in my glass count as an MSSC m-individual so long as any water that is connected to it counts as part of the water-whole.

If it turns out that WATER$^{\sqrt{c}}$ contains MSSC m-individuals, we have a problem. This is because WATER$^{\sqrt{c}}$ could combine with the IND filter, ultimately leading to singular count terms for WATER. This is undesirable, as it violates the mapping property where there are no known languages that have countable terms for liquids (see §4.2.1).

We should therefore want the IND filter to come out empty when combining with canonical $\sqrt{\text{SUB}}$ roots like WATER$^{\sqrt{c}}$. The simplest and easiest way to achieve this is to posit that there are no pre-theoretical individuals in a $\sqrt{\text{SUB}}$ root concept. To model this, there is a necessary discussion of the interaction between what is perceivable and what is linguistically accessible. In this manner, while we may perceive water as an MSSC m-individual at a construed ontological level, at the linguistic representational level there are no pre-theoretical individuals, MSSC or otherwise, in the root concept.

I model this intuition through the interaction of the three differing levels of ontology. The `FUNDAMENTAL ONTOLOGICAL` level is the ontology of the metaphysical, which concerns the structure of (m-individuals of) the world in an objective sense (if such a thing is possible). The `CONSTRUED ONTOLOGICAL` level is the ontology of human perception, which concerns how humans perceive and structure the world. This includes the pre/extra-linguistic representation of the substance/object distinction (see §4.1.2). The `LINGUISTIC ONTOLOGICAL` level is the ontology of natural language, and concerns the structure of linguistic items and how they refer. This is the level where root concepts sit.[15]

I take the levels of ontology to be ordered in such a way that higher level ontologies are built form lower level ontologies; each level is an abstracted representation of the level below it. The linguistic ontological level is the highest and is an abstracted representation of the construed ontological level which is in turn is an abstracted representation of fundamental ontological level. In the realms of the count/mass distinction, the interface of the construed ontological level and the linguistic ontological level is the root concept, where properties from the construed ontological level percolate to the linguistic ontological level. The basic idea is sketched in Figure 7.6.

[15]See also Bach (1986b); Kiss et al. (2021); Moltmann (2021); Pelletier (2011) for further discussions on the interactions between levels of ontology.

Fundamental ontologies	Entities in the world
	$\downarrow$
Construed ontologies	Construals of entities in the world
	$\downarrow$
Linguistic ontologies	Root concept

Figure 7.6: Fundamental, construed, and linguistic ontological mapping

The claim I wish to make is that the mapping from the construed ontological level to the linguistic ontological level (*i.e.* property to root concept) is multifaceted. I suggest that only properties with a construed INDIVIDUATION STANDARD, judged at the construed ontological level, percolate to linguistic ontological level to become an $\sqrt{\text{IND}}$ or $\sqrt{\text{SUB/IND}}$ root.[16] Here enters the discussion of FUNCTION and BOUNDARIES that hold over time and space.

The role of FUNCTION affords scattered wholes pre-theoretical individual status. This is demonstrated by the intuitive differences between the join of *bikini-top*$\sqcup$*bikini-bottom* and the join of *Klaus*$\sqcup$*the-Eiffel-Tower*. At the fundamental ontological level, these sums are akin (they are joins of non-connected MSSC m-individuals). However, only the bikini-sum is useful in a real human sense. This usefulness judgement is made at the construed ontological level. Therefore, while knowledge that bikinis are scattered wholes is present at the construed ontological level, so is the knowledge that they form a non-arbitrary sum that counts as the construed standard of individuation. It is precisely this judgement of function that affords a standard of individuation to *bikini-top*$\sqcup$*bikini-bottom* but not *Klaus*$\sqcup$*the-Eiffel-Tower*, allowing the bikini-m-individuals to percolate to the linguistic ontological level to be represented as a root concept.

[16]Taking individuals to have an individuation standard and liquids to not have an individuation standard is not controversial. Discussions for standards of individuation appear as early as Quine (1960). In more recent scholarly contexts, the discussion comes under terms of precisifications across contexts Chierchia (2010b) and judgements of disjointedness across all contexts Landman (2020).

The judgement of BOUNDARIES separates liquid and non liquid concepts. Consider the following variation on classical discussions of cumulativity, which I borrow from Casati and Varzi (1999, p.87-88)'s discussion of 'topological catastrophes', which I also discussed in §3.5. We may have two disjoint MSSC instances of oil, say two drops, and we then push them together such that the oil droplets merge. We have gone from having two MSSC oil m-individuals to having one MSSC oil m-individual. There has been absolute topological change of boundaries. Change of boundaries for liquids is easy. I can repeat the same experiment on water, milk and blood. The same cannot be done for solid m-individuals like ducks, computers, cars, or any solid item. Pushing two cats together does not result in one larger cat. In an intuitive sense, the boundaries of solid and liquid MSSC m-individuals are different; the boundaries of solid m-individuals hold over time and space without (easy) distortion or change, while the boundaries of liquids do not. With this, while liquids may have MSSC manifestations in a given time, space or context, and we may have knowledge of this, there is no individuating standard for liquids that holds across time, space or context. This is the difference.

This discussion of boundaries holding across time and space is clearly a modal notion. We are essentially discussing the potential for m-individuals to be an MSSC with respect to a property in one world/context, but a proper part of an MSSC m-individual of the same property in another world/context. The difference between pre-theoretical substances and pre-theoretical individuals is that for pre-theoretical individuals, there is no world/context where the m-individual can be a proper part of a m-individual of the same property. For pre-theoretical substances of the liquid variety, there is a world/context where that individual may be subsumed as a proper part of an m-individual of the same property. While I do not offer a formalisation, I will nonetheless rely on this intuition and assume (54) as the working definition of an INDIVIDUATION STANDARD.[17]

[17] Modelling individuation standards across contexts may take form as Chierchia (2010b)'s precisifications across contexts in a supervaluationist sense, as Landman (2020)'s overlap at every context, or through countermereology as in Casati and Varzi (1999).

(54) A property P has an `INDIVIDUATION STANDARD` iff any m-individual fulfils the properties:

a. is defined for INDIVIDUAL(x)

b. has persisting boundaries across time and space (in another world/context any m-individual relative to that property P cannot be a proper part of a m-individual of the same property).[18]

I propose (55) for the mapping of the construed ontological level to the linguistic ontological level, *i.e.* properties into root concepts.

(55) Mapping from properties to root concepts:

a. The individuation standard of a property, if it exists, is always mapped to the corresponding root concept. This individuation standard will always be compatible with the IND filter.

b. Firmly connected m-individuals that share a property are mapped to the root concept. These firmly connected m-individuals will always be compatible with the SUB filter.

The theory plays out like this. For solid non-scattered wholes with non-naïvely homogenous parts (*e.g.* dog), only the individuation standard is mapped to the root concept. Dogs are not firmly connected to other dogs, so there is no substance mapping. For liquid m-individuals (*e.g.* water) there is no individuating standard at the construed ontological level, and so no individuating standard, MSSC or otherwise, is mapped to the root concept at the linguistic ontological level. There is, however, a mapping of firmly connected m-individuals to the root concept. Finally, if any m-individual is construed as both having an individuation standard and comprising of constituent substance matter (*e.g.* stones), then both the firmly connected like-parts and the individuation standard are mapped into the root concept.

[18] There is a tension here with concepts such as FENCE and BOUQUET, which *e.g.* Rothstein (2010) points out are not bounded across time and space in the way defined here. How such nouns are incorporated into this theory remains to be seen.

Therefore, once entities are represented at the linguistic ontological level via root concepts, when the IND filter applies, it comes out empty for any given $\sqrt{\text{SUB}}$ root, but non-empty for $\sqrt{\text{IND}}$ root and $\sqrt{\text{SUB/IND}}$ root. In this manner, we may have knowledge or a judgement that an instance of a liquid in a given context fulfils a property of being MSSC relevant to a property, but this information is not encoded in the root concept. In this way, the linguistic ontological level of representation is abstracted; root concepts are not direct reflections of all possible construed ontological construals of a property, but representations of standards.[19]

A change of perspective

At the construed ontological level, there are entities which have both construed standards of individuation and construals of substancehood. In these cases of dual judgements, both the individuation standard and substance judgement percolate to the linguistic ontological, manifesting as a $\sqrt{\text{SUB/IND}}$ root. Ultimately, this leads to

[19]Yet, there is an exception to this in the case in which there are constant manifestations of water across time and space, *i.e.* lakes, waterfalls, ponds, oceans, seas. These are construals of permanent structures of water, where boundaries are vague only insomuch that a prototypical object is construed as vague (see Chierchia (2010b) for discussions of vagueness). Therefore, for terms like LAKE, though they denote liquids, they will be compatible with the IND filter. We can take this even further. I suggest that specifically the root $\text{WATER}^{\sqrt{c}}$ is actually compatible with the IND filter, but the only m-individuals that count as individuals that hold their boundaries over time and space. Such as geographical waters. If this is the case, we may expect WATER to behave as a count term only where it applies to constant bodies of water. This is exactly the interpretation of (ia). The first sentence is interpretable because WATER refers to a constant body of water, *i.e.* the seas and oceans that belong to England. Note, the issue of where boundaries of what counts as one body of water and what counts as another body of water is not a linguistic issue, but a mereotopological issue that is similar to the fiat boundary issue that separates *e.g.* France and Germany (see Casati and Varzi (1999) for more details). The interpretation of (ib) is grammatical, because there are constant bodies of Water in York, namely the rivers Ouse and Foss. Conversely, the examples in (ii) are deviant. There are no MSSC water-m-individuals constant across time of space either in my house nor in my bath-tub.

(i) a. The waters of England are very salty
b. The waters of York have been contaminated

(ii) a. ?The waters in my house are very salty
b. ?The waters in my bath have been contaminated

the potential for perspective shift, where either the IND filter or the SUB filter may apply.

I argue that perspective shift is precisely the source of nominal flexibility for fruit and vegetable terms. The idea is that fruit and vegetables construals have both standards of individuation and a judgement of substancehood, and both these judgements percolate to the root concept at the linguistic ontological level. This view requires there to be some construed ontological judgement concerning the material composition of fruits and vegetables such that it fulfils the requirement of being considered a substance, for which I rely on the definition of firm connectedness as in (53), repeated here for convenience.

(53) `SUBSTANCE`

$\textsc{substance}(\mathrm{x}) \stackrel{\text{def}}{=} \forall \mathrm{P}[\mathrm{P}(\mathrm{x}) \wedge \exists \mathrm{y}[\mathrm{P}(\mathrm{y}) \wedge \mathrm{x} \neq \mathrm{y} \wedge \textsc{fc}(\mathrm{x},\mathrm{y})]$

x is a substance if. for any property P x is firmly connected to a distinct m-individual of the same property.

From a fundamental ontological perspective, the view that individuals which comprise vegetation have substance constituent matter is incorrect: the pulp and skin of a tomato are not firmly connected m-individuals of the same property. However, at the construed ontological level, the idea that tomatoes are constituted of a tomato-substance is in some sense intuitively correct. The stuff that makes up tomato is in its own right tomato, even if there is no physical change applied to any given tomato. I submit this is a judgement that comes of function specific to fruit and vegetable terms: in normal day to day use, the parts of a tomato serve the same function as the whole: to be eaten. In this manner, the parts of a tomato are firmly connected to an m-individual of the same property, *i.e.* the property of tomato parts to be eaten, so to speak. This is not to say we cannot or do not distinguish between different constituent make-up of fruit and vegetable individuals. We most certainly can. Rather, on a functional level there is some judgement of fruit and vegetable individuals such that they consist of functional substance matter.

There is an obvious parallel here to the case of the bikini. While we know that the two parts of a bikini form a scattered whole, there is a judgement that the scattered whole is useful on some level. The case of fruits and vegetables is in the opposite direction: it is not sums of arbitrary wholes that are relevant, but the parts of the MSSC m-individuals that are afforded usefulness status, due to their constituent matter being edible foodstuff.

Note, a construal of edible function does not necessarily lead to a judgement that the constituent make-up is substance matter. Many animals are edible, and yet do not have construals that they comprise of substance matter. It is not intuitively correct to say that chickens are made up of chicken-stuff in the way that tomatoes are made up of tomato-stuff. To this, I suggest that the functional substance construal only arises if the majority of constituent matter of a given individual is judged as a functional substance in a majority of contexts. After all, even though chickens are commonly eaten, the majority of the parts and the whole of a chicken do not typically constitute food stuff, and so *e.g.* chickens are not construed as having a composition as functional substances, at least in their primary construals. As such, properties such as CHICKEN are represented only by $\sqrt{\text{IND}}$ concepts at the linguistic ontological level.

7.3.1.3 Putting this together

There is a fine line to tread here concerning fundamental ontologies and construed ontologies. Information at the fundamental ontological level is certainly available at the construed ontological level, but there is information at the construed ontological level that is beyond fundamental ontology (such as individuation standards, judgements of function). The view taken here is that root concepts are shaped by construed ontologies, not fundamental ontologies. Therefore, what is important is what is *construed* rather than what *is*. In the case of fruits and vegetables, there is a standard of individuation, but also a construed judgement where these fruit and vegetable individuals have substance composition. Therefore, when mapping fruit and vegetable properties from the construed ontological level to the linguistic ontological level, a root concept is created that is compatible with both the IND filter and the SUB filter. Putting together the ingredients for mapping different ontologies, the picture in (56) arises.

(56) a. There are properties at the construed ontological level that will map only to an $\sqrt{\text{IND}}$ root at the linguistic ontological level
b. There are properties at the construed ontological level that will map only to a $\sqrt{\text{SUB}}$ root at the linguistic ontological level
c. There are properties at the construed ontological level that will map to a $\sqrt{\text{SUB/IND}}$ root at the linguistic ontological level

An example of the property$\rightarrow\sqrt{\text{IND}}$ mapping is DUCK. At the construed ontological level, ducks have an individuating MSSC standard that holds across time and space, and there exists no duck m-individual which is firmly connected to another duck m-individual. The individuating standard (MSSC) percolates to the linguistic ontological level, creating an $\sqrt{\text{IND}}$ root where m-individuals are disjoint. As such, applying the IND filter will output a non-empty filtered set; applying the SUB filter will output an empty filtered set. A similar process applies to the property BIKINI, but in this case the individuating standard is a scattered whole.

An example of the property$\rightarrow\sqrt{\text{SUB}}$ mapping is WATER. At the construed ontological level, water does not have an individuating standard that holds across time and place, but all water m-individuals are firmly connected to another water m-individual. The firmly connected m-individuals percolate to the linguistic ontological level, creating a $\sqrt{\text{SUB}}$ root where m-individuals are vertically overlapping. As such, applying the IND filter will output an empty filtered set; applying the SUB filter will output a non-empty filtered set.

Finally, an example of the property$\rightarrow\sqrt{\text{SUB/IND}}$ mapping is STONE or CARROT. At the construed ontological level, there is an individuating standard that holds across time and place (MSSC), but at the same time the material composition of stone/carrot is made up of firmly connected m-individuals of the same property. All parts and wholes percolate to the linguistic ontology, and both the application of the IND filter and the SUB filter will result in a non-empty filtered set.

construed ontology		linguistic ontology
WATER(P), OIL(P)		WATER$^{\sqrt{c}}$
Substance ✔	$\longrightarrow$	$\sqrt{\text{SUB}}$
Individuating standard ✖		
DUCK(P), BIKINI(P)		CAT$^{\sqrt{c}}$
Substance ✖		
Individuating standard ✔	$\longrightarrow$	$\sqrt{\text{IND}}$
STONE(P), CARROT(P)		STONE$^{\sqrt{c}}$, CARROT$^{\sqrt{c}}$
Substance ✔	$\longrightarrow$	$\sqrt{\text{SUB/IND}}$
Individuating standard ✔	$\longrightarrow$	

Figure 7.7: Construed ontological to linguistic ontological mapping

7.3.2 GRIND

I now move on to account for stuff-readings of singular count and singulative terms. The stuff-readings of English DOG, LION and Welsh LEW (lion.SG = *'lion'*), MOCH-YN (pig.COL-SING = *'pig'*) in the repeated examples below are not due to perspective shift. On a conceptual level, dogs, lions and pigs are MSSC m-individuals with proper parts that do not constitute (functional) substances, and so the root concepts for these terms do not contain firmly connected m-individuals of the same property.

(4a) There is **dog** all over the wall — English

(4b) There is **lion** in my soup — English

(33a) Mae yna **lew** yn y cawl — Welsh
be.3 there lion in DEF soup
Stuff: '*There is lion meat in the soup*'
Integral: '*There is a lion in the soup*'

(37b) Mae yna **moch-yn** yn y cawl — Welsh
be.3 there pig-SING in the soup
Stuff: '*There is pig meat in the soup*'
Integral: '*There's a pig in the soup*'

I submit that the stuff-readings in these cases arise from a COUNT → MASS device, GRIND, in a Rothstein (2010, 2017) inspired manner. I argue that not only singular count terms may be the input to GRIND, but so too can singulative terms. The uniting feature of these terms is semantic singularity, ⟨**body**, **base**⟩, serving as the input to GRIND, as schematically represented in (57).[20]

(57) Schematic 1 for GRIND
GRIND = ⟨**body**, **base**⟩, where the **base** is disjoint (count) →
⟨**body**, **base**⟩, where the **base** is overlapping constituent stuff (stuff mass)

Unlike Rothstein (2010, 2017), I do not take GRIND to necessarily be a last resort device. This is partially due to the free application of GRIND in Welsh, but also its application to plural marked terms in English. Consider the examples (58)-(60). There is a stark difference between DOG/CHICKEN and NUT/WORM/FLY/PRAWN with respect to available stuff-readings. For DOG, the lack of number marking results in a stuff-reading. However, the lack of number marking for NUT/WORM/FLY/PRAWN is awkward in these contexts, and the plural marked term is used for *both* the plural (integral) reading and the stuff-readings. For example,

[20] There is perhaps an alternative analysis where MOCH-YN is not a singulative marked term that has undergone grinding, but -YN is a partition/portioner that applies only to collective terms. I do not pursue this analysis, but incorporate the stuff-readings of MOCH-YN as reflective of functions that are independently motived. That is, as singulative markers are semantically singular, and semantically singular animal terms can gain stuff-readings, there is no need to propose alternative mechanisms for MOCHYN.

(58b) is compatible with a situation where there are multiple whole nuts floating in my smoothie, perhaps added after the initial blending (integral-reading), but also a situation where the nuts were blended up as part of the smoothie (stuff-reading). The effect is seen not only with canonical food stuff. The same ambiguity arises with predicates denoting individuals that are not typically eaten, such as WORM and FLY; example (58b) is compatible with whole worms wriggling in my smoothie, or worms having been blended into my smoothie. As far as I can tell, the preferential plural marking for stuff-readings seems to apply to terms which naturally come in groups and are small.

(58) a. There's **dog/chicken/alpaca** in my smoothie — English
b. There's **nuts/worms/flies/prawns** in my smoothie
c. #There's **nut/worm/fly/prawn** in my smoothie

(59) a. This product may contain **dog/chicken/alpaca** — English
b. This product may contain **nuts/worms/flies/prawns**
c. #This product may contain **nut/worm/fly/prawn**

(60) a. I'm allergic to **dog/chicken/alpaca** — English
b. I'm allergic to **nuts/worms/flies/prawns**
c. #I'm allergic to **nut/worm/fly/prawn**

I suggest that the difference in felicitousness of singular/plural stuff-readings (in at least English) is a matter of context/pragmatic choice: if multiple entities are required to get a substantial amount of 'stuff' then plural marked terms are the preferential input to GRIND. The major consequence that falls out from this proposal is that GRIND is not always a last resort mechanism triggered by incongruent morphosyntax. This is not to say that GRIND is never invoked for last resort. Rather, last resort is not the only trigger for GRIND. Other triggers of GRIND seemingly include contexts that prejudice stuff mass readings. We therefore update the schematic for GRIND as in (61).

(61) Schematic 2 for GRIND
GRIND = [⟨**body**, **base**⟩, where the **base** is disjoint (count)] OR [⟨***body**, **base**⟩, where the **base** is disjoint (plural count)] → ⟨**body**, **base**⟩, where the **base** is overlapping constituent stuff (stuff mass)

I now formalise the schematic in (61). To do this, a function that reveals parts of m-individuals is required. This is captured by the constituent parts function, Ω(P), in (62) which takes all the elements of a given set and returns a set of all the proper parts of each element. If applying Ω(P) to a disjoint set that contains three ducks, the output of this function will be a vertically overlapping set of duck-parts (*e.g.duck-bill*$\sqcup$*duck-tail* and *duck-bill*$\sqcup$*duck-tail*$\sqcup$*duck-wing etc.*.) In this manner, all the parts of the original MSSC ducks are present in the set, but there is no MSSC configuration of these parts with respect to the property DUCK.

(62) CONSTITUENT PARTS
$\Omega(P) \rightarrow \lambda P \lambda y \forall x[P(x) \rightarrow (y \sqsubset x)]$
Ω(P) is a function such that for any set P and any x in P, Ω maps x onto the set of its proper parts.

We now take Ω(P) and apply it to the **body** of an i-set. Assuming BODY(P) and BASE(P) in (63) and (64), then the semantics for GRIND are as in (65). This function takes an i-set with a disjoint **base**, extracts the **body** of that i-set, and then maps that **body** onto the set of the constituent parts of the individuals in the **body**, and then uses that set as the **body** and **base** of a resulting i-set. As the **base** of the new i-set is vertically overlapping *stuff*, the new i-set is a stuff mass.

(63) BODY(P) is a function that maps any i-set P to the set of its **base**.

(64) BASE(P) is a function that maps any i-set P to the set of its **base**.

(65) $$\textsc{grind}(\mathrm{P}) \rightarrow \lambda \mathrm{P}.\begin{cases} \langle \Omega_{(\textsc{body}(\mathrm{P}))},\ \Omega_{(\textsc{body}(\mathrm{P}))} \rangle & \text{if } \textsc{base}(\mathrm{P}) \text{ is disjoint} \\ \perp & \text{otherwise} \end{cases}$$

The semantics proposed for GRIND are indifferent as to whether a semantically singular count or a plural count i-set may function as input. This is modelled as requiring the input i-set to have a disjoint **base**. This captures that English count or plural terms may be input to GRIND, but also that Welsh singulative i-sets may be input to GRIND. There is also the consequence that number neutral uncountable terms are incompatible with GRIND, due to the fact they have an overlapping **base**. This is desirable, as collective terms for meaty/fleshy animals do not gain stuff-readings easily in English, Welsh or Mandarin Chinese.

With this, note that GRIND will always be accompanied by notions of physical change of individuals. This is because GRIND will only ever apply to an i-set that is built via the IND filter. In this manner, the pre-theoretical individuals serve as input to the function which reveals their constituent parts.

Even with this new understanding for the nature of GRIND for stuff-readings for singular count, singulative, and plural terms, the crosslinguistic implementation of GRIND is not clear. In English, GRIND seems to be not triggered by last resort, but is trivially free in where the preferential input (singular/plural) is modulated by context/contextual factors. In Welsh, GRIND is not restricted to last resort, and may apply whenever it is contextually compatible. In Semitic languages, such as Arabic and Modern Hebrew, it is not clear that GRIND ever applies.[21] The freeness of GRIND therefore cannot be universal stated. Rather, the only universal that is possible to suggest is that GRIND does not apply to number neutral uncountable terms as ⟨***body**, ***base**⟩ i-sets are not semantically appropriate for the input.

[21] Unless triggered by a gender mismatch in Modern Hebrew, see Rothstein (2017).

(66) CLAIM:
Stuff readings of count and singulative ⟨***body**, ***base**⟩ and plural ⟨***body**, **base**⟩ i-sets are derived via GRIND.

a. The availability of GRIND across languages varies.

b. The use of GRIND is modulated by (language specific) contextual requirements.

c. GRIND will never apply to ⟨***body**, ***base**⟩ i-sets, and so number neutral uncountable terms are resistant to stuff-readings.

7.3.3 Accidentally homophonous root concepts

I have so far proposed that all fruit and vegetable terms may be used in a stuff mass way due to construals of vegetation as a functional substance at the construed ontological level, a feature which percolates to the root concept at the linguistic ontological level (perspective shift). Following this, I proposed GRIND to account for stuff mass use of canonical count, singulative and plural terms in English and Welsh, which may be triggered by context. From this, number neutral uncountable terms which denote individuals which do not have construals of constituent substance matter should resist stuff-readings. For the most part, this prediction is borne out: non-vegetable number neutral uncountable terms (including collectives) are typically resistant to stuff-readings in Welsh and Mandarin Chinese. However, this does not explain the exception cases seen in Hejazi Arabic where SAMAK (fish.COL = *'fish(es)'*) and DEJAAJ (chicken.COL = *'chicken'*) may gain stuff-readings in amenable contexts, as in (40), repeated. These stuff-readings should be ruled out as *e.g.* chickens are not composed of naïve substance matter, nor are these terms valid inputs to GRIND, due to the overlapping **base** of number neutral uncountable terms.

(40) b. fi **dejaaj** fi shurbat-ii — H. Arabic
in chicken in soup-my
Integral: '*There are chicken/s in the soup*'
Stuff: '*There is chicken meat in the soup*'

c. fi **samak** fi shurbat-ii — H. Arabic
in fish in soup-my
Integral: '*There are fish/es in the soup*'
Stuff: '*There is fish meat in the soup*'

To explain the interpretation of DEJAAJ and SAMAK as stuff mass terms, I take a seemingly sharp detour into English terms for the meat of animals. In English, there are stuff mass terms for specific animal meats including PORK, GAMMON, BACON, HAM (meat from pigs), MUTTON, LAMB (meat from sheep), BEEF, VEAL (meat from cows) and VENISON (meat from deer). Being stuff mass, specific meat terms are predictably felicitous in contexts that favour stuff-readings as in (67a). Notice however, that in the same contexts the count terms for the animals from which the meat comes from are extremely awkward (67b), as is the use of a compound (67c). This awkwardness does not hold for count terms for animals which do not have a corresponding specific stuff mass term for the animal's flesh, whether the animal is commonly eaten or not, as in (68). Following Cheng et al. (2008)'s wisdom, I suggest there is a blocking mechanism, PREFSUB, as in (69).

(67) a. There's **beef/mutton/pork** in my soup — English
b. # There's **cow/sheep/pig** in my soup
c. # There's **cow/sheep/pig meat** in my soup

(68) a. There's **salmon/rabbit/chicken/duck/dog/lion** in my soup — English
b. There's **salmon/rabbit/chicken/duck/dog/lion meat** in my soup

(69) PREFER SUBPREDICATIVE STRUCTURES (PREFSUB)
If there is a choice between a subpredicative stuff mass term composition, or an i-set with further structure (GRIND, compound), prefer the subpredicative composition.

PrefSub assumes there are multiple structures that are able to derive stuff-readings, and they are all in competition. For this to work, the stuff mass derivation [BACON$^{\sqrt{c}}$ + SUB] is preferred over a [PIG$^{\sqrt{c}}$ + IND + GRIND] complex. How can this be modelled?

First, consider that at the construed ontological level, MSSC m-individuals may have proper parts. It may very well be the case that those proper parts are themselves conceptually relevant in some sense, perhaps even MSSC m-individuals themselves relative to a *different* property than the whole it is part of. For example, my cat Klaus is an MSSC m-individual relative to the property of CAT, but each of his paws are MSSC m-individuals relative to the property of PAW. This does not deny that the paws are part of Klaus. Rather, MSSC status is only ever relative to a certain property. Therefore, even though there is a clear part-whole structure between Klaus and his paws, this is differentiated by different properties at the construed ontological level, which are then lexicalised at the linguistic ontology level as specific root concepts that form a part-whole pair: CAT$^{\sqrt{c}}$ and PAW$^{\sqrt{c}}$.

By analogy, we can apply this reasoning to terms for animals and their meats. Firmly connected beef m-individuals are proper parts of MSSC cow m-individuals. Firmly connected bacon m-individuals are proper parts of MSSC pig m-individuals. It just so happens that there is a representation of the firmly connected meat-parts at the root concept level, but unlike paw individuals, the firmly connected meat m-individuals are (functional) substances, and so represented as $\sqrt{\text{SUB}}$ roots. In this manner, the root concepts BACON$^{\sqrt{c}}$ and PIG$^{\sqrt{c}}$ are related via their part-whole relationships at the construed ontological. This relationship underlies the very essence of PrefSub: the meat-stuff of an *e.g.* a pig can be linguistically represented via multiple means, *i.e.* there is a stuff mass derivation via [BACON$^{\sqrt{c}}$ + SUB], but also a compound derivation via [PIG$^{\sqrt{c}}$ + IND + compound] and a GRIND derivation via [PIG$^{\sqrt{c}}$ + IND + GRIND]. All of these methods are in competition, but the stuff mass derivation is favoured by PrefSub.

Typically, terms for related construed ontological properties are phonologically distinct in English: BACON/PIG, BEEF/COW, MUTTON/SHEEP. However, this need not always be the case. Consider English LAMB, which has clear integral-reading (70), but also a stuff-reading where the meat is referred to (71). I submit that LAMB is an example of what I call an `ACCIDENTALLY HOMOPHONOUS ROOT CONCEPT`, where the root concept for the MSSC m-individual (the animal) and the root concept for the firmly connected meat-stuff of that animal are distinct but have the same phonological realisation after predicate composition. In this way, LAMB is represented a la BACON/PIG as two related root concepts. This being the case, accidentally homophonous root concepts do not represent a stuff-reading reading derived either by GRIND nor perspective shift. Rather, it is simply two separate structures of related concepts that happen to have the same phonological realisation.

(70) There are **lambs** in the garden English

(71) There is **lamb** in my soup English

But, is there any proof for accidentally homophonous root concepts? If there are two roots for LAMB, a $\sqrt{\text{SUB}}$ root and an $\sqrt{\text{IND}}$ root, then we may expect PREFSUB to apply and the [$\sqrt{\text{SUB}}$ + SUB] composition should be preferred over the [$\sqrt{\text{IND}}$ + IND + compound] structure for reference to the meat matter. Therefore, LAMB MEAT should be blocked in the way that PIG MEAT is blocked. As it turns out, this is the case: (72) is extremely awkward.

(72) # There is **lamb meat** in my soup English

Now we return to Hejazi Arabic. I propose that the ambiguity of integral-reading and stuff-readings of collective terms DEJAAJ and SAMAK in (40) are cases of accidentally homophonous root concepts. The evidence for this comes from the fact that these terms are extremely awkward in compound structures as in (41). This indicates that, like English LAMB, Arabic DEJAAJ and SAMAK have competing derivations for the stuff-readings: the stuff mass composition [DEJAAJ$^{\sqrt{c}}$ + SUB], and the compound derivation [$\sqrt{\text{IND}}$ + IND + **f:set* + compound]. This materialises as an ambiguity in stuff mass and integral composition as in (40), and the compound derivation in (41) being dispreferred/blocked by PREFSUB.

(40) b. fi **dejaaj** fi shurbat-ii — H. Arabic
in chicken in soup-my
Integral: '*There are chicken/s in the soup*'
Stuff: '*There is chicken meat in the soup*'

c. fi **samak** fi shurbat-ii — H. Arabic
in fish in soup-my
Integral: '*There are fish/es in the soup*'
Stuff: '*There is fish meat in the soup*'

(41) a. #fi **lahm dejaaj** fi shurbat-ii — H. Arabic
in meat chicken in soup-my
Intended: '*There is chicken meat in my soup*'

b. #fi **lahm samak** fi shurbat-ii — H. Arabic
in meat fish in soup-my
Intended: '*There is chicken meat in my soup*'

Further support for accidentally homophonous root concepts in Hejazi Arabic comes from the behaviour of collective term BAQAR (cow.COL = *'cows'*) (these facts also hot for BAṬ-AH (duck.COL-SING = *'duck'*)). As a term for a meaty/fleshy animal, we expect an $\sqrt{\text{IND}}$ root concept. As a collective term, BAQAR is a ⟨***body**, ***base**⟩ i-set. Therefore, we would expect BAQAR to resist stuff-readings, as neither perspective shift nor GRIND may apply. This is the case: (40a), repeated below, is not compatible with a stuff-readings. Now, as there is no ambiguity in the reading of (40a), then is must be the case there is no accidentally homophonous root concepts; there is no separate $\sqrt{\text{SUB}}$ root concept for the meat-stuff of cows that is realised as BAQAR. If this is true, then PREFSUB does not apply, and compounds are used with the collective term for stuff-readings. This turns out to be the case, as in (73).

(40a) fi **baqar** fi shurbat-ii H. Arabic
in cow in soup-my
Integral-reading only: '*There are cow(s) in the soup*'

(73) fi **lahm baqar** fi shurbat-ii H. Arabic
in meat cow in soup-my
'*There is cow meat in my soup*'

7.4 Discussion

The system presented rejects the notion that there is a single universal grinder that applies to all nouns in all languages which is responsible for stuff-readings. Rather, the source of stuff-readings is argued to be a multi source, including at the very least: perspective shift, GRIND, compounding, and accidentally homophonous root concepts. Modulating between the available methods in and across languages is intricate, with very specific candidates for universal processes.

I have proposed that `PERSPECTIVE SHIFT` is a universal method of gaining stuff-readings in and across languages. The basic idea is that if an individual has a construed standard of individuation *and* its material composition is construed as a substance, then the corresponding linguistic representation will be a $\sqrt{\text{SUB/IND}}$ concept which may compose in either a stuff mass or an integral (singular count, number neutral uncountable) way. In a sense, perspective shift is the successor of constructionist structural ambiguity for stuff-readings in the Borer (2005a); Bale and Barner (2009) way. The key difference is that in this Subpredicative Iceberg Semantics rendition of structural ambiguity, there is no universal structural ambiguity for all terms. Only root concepts which contain information pertaining to substance and individual status may be systematically ambiguous. The idea is summarised in (74).

(74) `PERSPECTIVE SHIFT IS UNIVERSAL`

If a property has a construed standard of individuation and the constituent make-up of an individual is a construed substance (naïvely homogenous), then that property may be mapped to a $\sqrt{\text{SUB/IND}}$ root concept, and can therefore compose in an integral way (count, number neutral uncountable) or stuff mass way.

a. Fruit and vegetable individuals are construed as comprising of functional substance matter, and are predicted to be universally useable in both integral and stuff mass ways.

For non ambiguous roots, *i.e.* an $\sqrt{\text{IND}}$ root which only contains individuals that have construed standards of individuation, alternative methods must be used for stuff-readings. This may come in the form of a `COMPOUND`, which is taken to be a universal method for deriving stuff-readings, or the use of GRIND, a COUNT → MASS shifting device inspired by the lexicalist last resort approach in the Cheng et al. (2008); Rothstein (2010, 2017) tradition. Yet, unlike Rothstein; Cheng et al., I do not take GRIND to necessarily be a last resort mechanism; I allow GRIND to be compatible with any countable term (*i.e.* i-sets with a disjoint **base**). I also take GRIND to be language specific: while English and Welsh have a GRIND operation, Hebrew and Hejazi Arabic have at best a highly restricted GRIND. With this, I propose (75).

(75) `COUNT → MASS SHIFTING DEVICES ARE LANGUAGE SPECIFIC`

a. Singular count and plural terms may gain stuff-readings via the COUNT → MASS shifting device GRIND.
b. Languages may differ on their permissible uses of GRIND. It may be a last resort mechanism (Hebrew, trivially Hejazi Arabic), or contextually licenced (English, Welsh).
c. Default nominal number composition affects availability of stuff-readings via GRIND. If a language has no singular count terms (Mandarin Chinese), then GRIND will never apply.

Finally, the case of `ACCIDENTALLY HOMOPHONOUS ROOT CONCEPTS` is curious. The proposal is that are two distinct root concepts, a $\sqrt{\text{SUB}}$ root and an $\sqrt{\text{IND}}$ root, which are homophonous but are conceptually related in that the physical entities associated with these roots form a part-whole relationship at the construed ontological level. I submit that cases of accidentally homophonous root concepts are rare; the *accidentally* part is key here, as most root concepts relating to part-whole relationships have distinct phonological realisations. The consequences of this end up being quite similar to perspective shift: there are two compositions, a stuff mass composition and an integral composition (singular count, number neutral uncountable) composition, which have identical phonological realisations. The difference between perspective shift and accidentally homophonous root concepts is that in the former there is one $\sqrt{\text{SUB/IND}}$ root concept, and in the latter there are two root concepts, $\sqrt{\text{IND}}$ root and $\sqrt{\text{SUB}}$ root. The idea is summarised in (76).

(76) `ACCIDENTALLY HOMOPHONOUS ROOT CONCEPTS`
When two distinct root concepts a.) have a part-whole relationship at the construed ontological level, and b.) share a phonological realisation, this is accidentally homophonous root concepts

a. accidentally homophonous root concepts are subject to PrefSub, and so integral derivations combining with compounds may be degraded.

Taking both (74) and (75) together, it emerges that number neutral uncountable terms which do not denote fruits and vegetables resist stuff-readings because neither perspective shift nor GRIND may apply. Therefore, number neutral uncountable terms must resort to *e.g.* compounding to gain stuff-readings. This has the consequence that bare animal denoting general number terms in Mandarin Chinese and animal denoting collectives in Welsh and Hejazi Arabic are resistant to stuff-readings. It follows, then, that any cases of stuff-readings for number neutral uncountable terms are illusionary, and can instead be analysed as cases of either perspective shift or accidentally homophonous root concepts.

(77) NUMBER NEUTRAL UNCOUNTABLE TERMS RESIST STUFF-READINGS

a. Number neutral uncountable terms resist stuff-readings, as the $\sqrt{\text{IND}}$ concept was never a candidate for the SUB filter, and the ⟨***body**, ***base**⟩ term is not a candidate for GRIND. Alternative methods, such as compounding, are used.

b. So called stuff-readings of fruit and vegetable number neutral uncountable terms are illusionary; the stuff-readings are via perspective shift, and there is homophony between the stuff mass and integral compositions.

c. So called stuff-readings of animal number neutral uncountable terms are illusionary; the stuff-readings are via accidentally homophonous root concepts where meanings and uses are modulated by PREFSUB.

Before closing this chapter, I briefly discuss an interesting consequence of the multiple source account for stuff-readings. Consider that in theory GRIND may apply to singular count or singulative terms that refer to fruits and vegetables. If PREFSUB is correct, singulative marked fruit and vegetable terms should be awkward, as the [$\sqrt{\text{SUB}}$ + the SUB filter] composition should be preferred over the [$\sqrt{\text{IND}}$ + IND + **f:set* + singulative + GRIND] composition. This seems to be the case for Welsh where the collective term is highly preferred for stuff-readings, as in (78). Insofar as I can tell, the stuff-readings for singulative term MORONEN in (78) is quite marginal, and if interpreted as a stuff-reading at all, there is a tendency to interpret as a single salient carrot in context has been chopped up and put in the soup.

(37d) Mae yna **foron** yn y cawl — Welsh
be.3 there carrot in DEF soup
Integral: '*There are carrots(s) in the soup*'
Stuff: '*There are chopped/blended carrots in the soup*'

(78) Mae yna **foron-en** yn y cawl — Welsh
be.3 there carrot-SING in DEF soup
Integral: '*There are carrots(s) in the soup*'
\# Stuff: '*There is a chopped/blended carrot in the soup*'

The situation in English seems to be more complicated. In English, PREFSUB apparently does not apply (or fails to apply) in some cases. Consider the examples in (79). As blueberries are fruits, the property BLUEBERRY should map to a $\sqrt{\text{SUB/IND}}$ root. Therefore, there should be a [$\sqrt{\text{SUB}}$ + SUB] composition which is preferred over the [$\sqrt{\text{IND}}$ + IND + plural + GRIND] composition as per PREFSUB. Yet, this is not always the case. There is a tension where for some speakers, the plural marked BLUEBERRIES in (79) is the preferable form.

(79) a. There's **blueberries** in my smoothie — English
b. There's **blueberry** in my smoothie

I tentatively suggest that the preference some speakers have for [$\sqrt{\text{IND}}$ + IND + plural + GRIND] in (79) is due to a pragmatic size effect. The intuition is that relatively small individuals would not provide much *stuff* in context unless they are in abundance/multiplicities. In these cases, [$\sqrt{\text{IND}}$ + IND + plural + GRIND] provides a multiplicity of individuals for grinding. This idea has support from examples such as (80), where larger fruits and vegetables become increasingly felicitous when morphologically unmarked. I take this as an indication that the multiplicity requirement of small individuals may, for some speakers, override PREFSUB in context.

(80) There's ______ in my smoothie English
Not plural marked:
#BLUEBERRY, #RASPBERRY, #BLACKBERRY, APPLE, PEAR, CARROT, POTATO, MANGO, COCONUT, PINEAPPLE, WATERMELON
Plural marked:
BLUEBERRIES, RASPBERRIES, BLACKBERRIES GRAPES, APPLES, PEARS, CARROTS, POTATOES, *MANGOS, *COCONUTS, *PINEAPPLES, *WATERMELONS

In all, though I have rejected the notion of a single universal grinder and instead offered a multi-source approach to the availability of stuff-readings, the outlook is still not entirely clear. It may turn out that some of the proposed methods here are (not) truly universal across languages. For example, perspective shift is assumed universal, but only further investigation will corroborate or deny this claim. Further, the nature of GRIND needs serious testing in and across languages; I have so far proposed a basic semantics, but have offered little discussion on how it is constrained. What is clear, however, is that the collective/singulative system of Welsh and Hejazi Arabic, and even the closer look at English, show that there is a delicate balance to be found that ties together percept, grammar, and context, and only when all three are considered in tandem can we approach understanding of how stuff-readings verses integral-readings are derived in and across languages.

7.5 Summary and next steps

This chapter has been dedicated to accounting for nominal flexibility in and across languages. The Subpredicative Iceberg Semantics account proposed here rejects the notion of a universal grinder and instead offers a multiple-source account for nominal flexibility, with perspective shift being a universal method, while GRIND is language specific.

There is much work to be done on investigating the availability of integral and stuff-readings of nominal terms in and across languages. The journey throughout this chapter has shown that the issue is wrought with intricacies and delicate judgements. What has emerged from the investigation of the so-called universal grinder from the point of view of singulative languages is that a.) stuff-readings cannot uniformly be a last resort mechanism, and b.) number neutral uncountable terms which denote fleshy animals seem to have a universal ban on gaining stuff-readings.

It may well turn out that the solutions offered in this chapter need fine-tuning. The hope is that this chapter has opened the field of enquiry into an investigation of stuff-readings from a multi-source approach rather than a simple universal grinder. Going forward, there is a question as to how applicable these conclusions are to other languages. The results here make very strong predictions regarding fruit and vegetable terms. The claim is that fruit and vegetable properties will systematically create $\sqrt{\textsc{sub/ind}}$ roots across languages due to construals of functional substancehood of the material of matter, and so should have systematic stuff mass uses across languages, subject to constraints such as PrefSub. All other sources for stuff-readings are language specific in nature. Another question we may ask is if PrefSub is correct, then why is CARROT PULP acceptable if CARROT can be used in a stuff mass way? I leave further investigation into the validity and expansion of these claims for future research.

Chapter 8

Summary and closing remarks

This thesis has explored disparate number systems, specifically the Welsh and Arabic collective/singulative systems from the point of view of the count/mass distinction. The specific research questions I set out to explore are repeated in R1 and R2.

R1 What are the primitive features that underlay collective/singulative, object mass, and general number systems, and what features are unique to each system?

R2 What is the architecture of nominal number?

a. Is it constructionist in nature, lexicalist in nature, or are these categories too restrictive?
b. What are the unifying semantics/syntax of object mass, collective and general number terms?
c. What are the unifying semantics/syntax of singulatives and (general) classifiers?
d. Why can the Arabic singulative pluralise, but the Welsh singulative cannot?
e. What can singulative languages tell us about the nature of nominal flexibility?

R1: Primitive features

In §2.2, I have shown that object mass, collective and general number are at their core number neutral uncountable terms in their descriptive behaviours: they do not directly combine with numerals, they are number neutral, and they have ontological commitments to (sums of) pre-existing individuals (and so are compatible with `STUBS` and comparison constructions).

So-called singulative markers are shown to have two distinct functions. Individuating singulatives were shown to be akin to the generalised classifier GÈ in Mandarin Chinese, where the unifying function derives a semantically singular term. Importantly, the individuating singulative and generalised classifiers preserve pre-existing ontological commitments to individuals. On the other hand, packaging singulatives were shown to be packaging/portioning devices which do not preserve pre-existing ontological commitments, but create individuals in context.

Though the grammatical behaviour of object mass, collective and general number are functionally comparable in terms of grammatical number, the distribution of number neutral uncountable terms in natural languages varies. Singulative languages afford to us the knowledge that the distribution of number neutral uncountable terms can be sensitive to perceptual realities in some, but not all, languages. While all object denoting terms in Mandarin Chinese are number neutral uncountable, only those which the denoted individuals which fulfil a sense of construed aggregation become number neutral uncountable in number marking languages and singulative languages (§4.2.3). Further, I proposed that some languages have secondary sensitivities that affect number neutral uncountable coding. Animacy, interaction, and even a minor borrowing effect are assume to be variable relevant to all languages, while size distinctions are proposed to be relevant specifically for Welsh (§5.2.2).

R2: architecture of nominal number

The interaction between the construed ontologies and the linguistic ontologies is multifaceted and complicated. It is clear that a theory of number cannot ignore the relevance of construed ontological reality for nominal structure. In particular, matters of construed aggregation and the substance/object distinction are key.

R2a. In chapter 4, I explored how theories of the count/mass typically either forgo crosslinguistic parsimony (the lexicalist approach) or perceptual facts (the constructionist approach) in the modelling of nominal number. This state of affairs is undesirable. To forgo crosslinguistic parsimony does not capture that object mass, collective and general number systems share the same primitive structures. To forgo perceptual facts does not capture that construed aggregation is intrinsically connected to collective/singulative systems. This thesis submits that while the linguist should strive to capture the connection between construed ontologies and linguistic ontologies, it should be done in such a way that the primitive features across all number systems are robustly captured. To capture both these desirable traits, Subpredicative Iceberg Semantics was suggested as a fruitful approach as the account aims to sacrifice neither perceptual facts nor crosslinguistic harmony.

R2b. In chapter 5, I proposed an extension to de Vries and Tsoulas (2021)'s Subpredicative Iceberg Semantics, which is a hybrid constructionist/lexicalist approach to nominal number. In the constructionist tradition, the locus of crosslinguistic variation in terms of the number neutral uncountable/count distinction was identified as a structural component.

In (extended) Subpredicative Iceberg Semantics, crosslinguistic variation of count/number neutral uncountable reference occurs during i-set derivation, where projection of **f:set* ensures number neutral uncountable composition ($\langle {}^*\textbf{body}, {}^*\textbf{base}\rangle$), and lack of **f:set* leads to singular count ($\langle \textbf{body}, \textbf{base}\rangle$) composition. The novel contribution of this thesis is that languages are necessarily either `FREE`* or `PRESUP`* with respect to **f:set* projection. `FREE`* languages, like Mandarin Chinese, obligatorily projected **f:set* for all individual denoting nominal compositions, resulting in blanket number neutral uncountable composition for nominal terms. `PRESUP`* languages, like English, Welsh, and Arabic, may associate **f:set* projection with a set of presuppositions. Following lexicalist inspiration, specifically Grimm (2012b), the set of presuppositions afforded to **f:set* concern construed aggregation of natural entities.

In this manner, number neutral uncountable terms in number marking languages, singulative languages, and generalised classifier languages are syntactically and semantically of the same type. The difference lies only in a language's ability to project the function which derives these terms, namely **f:set*. Importantly, the projection of **f:set*, while influenced by perceptual realities of speakers in `PRESUP`* languages, is inherently linguistic. This is because speakers of different languages may well construe the world in the same ways, but languages differ in the presuppositions associated with **f:set* in predictable ways. Therefore, the extended Subpredicative Iceberg Semantics approach maintains crosslinguistic parsimony of the constructionist approach, while at the same time incorporating the lexicalist wisdom of construed ontological modelling.

R2c. In chapter 6 Individuating singulatives and general classifiers are given a uniform semantics: $\textsc{cl}_{\textsc{sing}}(\mathrm{P})$. This function derives a singular count i-set from a number neutral uncountable i-set ($\langle {}^*\mathbf{body}, {}^*\mathbf{base}\rangle \rightarrow \langle \mathbf{body}, \mathbf{base}\rangle$). The semantics of $\textsc{cl}_{\textsc{sing}}(\mathrm{P})$ are considered to be one the the most basic functions of number, and so do not require combination with number neutral uncountable which denote certain perceptual features (such as clusters a la Grimm (2012b)). In this manner, the individuating singulative marker itself is purely linguistic, and does not reflect any construed ontological construals. This does not rule out that certain classifiers may have specific perceptual requirements (*i.e.* specific classifiers), but that singulatives and general classifiers are not of this kind.

R2d. Understanding why the Welsh singulative does not pluralise requires understanding the nature of the Welsh plural. In 2.2.2, I demonstrated that the Welsh overt plural is distinct from the plural in English and the plural in Arabic, as it cannot combine with numerals. Taking the observational facts at face value, and assuming the inclusive approach to plurality, a simple and intuitive analysis emerges. In chapter 6, I proposed that English and Arabic overt plurals have $\textsc{pl}_{\textsc{c}}(\mathrm{P})$

semantics, *i.e.* a function which take a singular count i-set and derive a countable plural i-set ($\langle {}^{*}\mathbf{body}, \mathbf{base}\rangle \rightarrow \langle {}^{*}\mathbf{body}, \mathbf{base}\rangle$). On the other hand, the Welsh overt plural has $\textsc{pl}_{\textsc{nc}}$(P) semantics, *i.e.* a function which takes a singular count i-set and derives an uncountable plural i-set ($\langle \mathbf{body}, \mathbf{base}\rangle \rightarrow \langle {}^{*}\mathbf{body}, {}^{*}\mathbf{base}\rangle$). With this, combining plural markers with singulative terms in Welsh and Arabic have different effects. Singulative marked terms in both Welsh and Arabic are valid inputs to their respective plural functions, but only in Welsh does the plural function applied to a singulative marker vacuous, *i.e.* in Welsh an i-set is derived which is semantically equivalent to the collective ($\langle {}^{*}\mathbf{body}, {}^{*}\mathbf{base}\rangle \rightarrow \langle \mathbf{body}, \mathbf{base}\rangle \rightarrow \langle {}^{*}\mathbf{body}, {}^{*}\mathbf{base}\rangle$). Such a derivation is ruled out as vacuous under Economy.

R2e. In chapter 7, I discuss the stuff-reading and integral-reading of collective/singulative terms and show that nominal flexibility cannot be purely constructionist of lexicalist in nature. First, the constructionist approach is ruled out as there is no complete nominal flexibility of unmarked terms: fleshy animal collective terms simply do not uniformly experience stuff-readings, though collective terms for fruits and vegetables do. Further, projection of 'division' cannot explain why Welsh singulative marked terms may also experience stuff-readings. On the other hand, nominal flexibility cannot be analysed as a purely lexicalist phenomenon in the manner of a last resort mechanism in cases of morphosyntactic mismatches, either. This is because Welsh systematically allows stuff-readings for semantically singular count (singular or singulative) in contexts that favour such readings, regardless of grammatical (in)congruity.

A multi-source approach to flexibility in the style of Subpredicative Iceberg Semantics was proposed. Taking inspiration for constructionist accounts, `PERSPECTIVE SHIFT` is the manner in which stuff-readings and integral readings are structural. Perspective shift is conceptual in nature and is not entirely free; it relies on the interaction between construed ontologies and linguistic ontologies. Specifically, individuals which have a naïvely homogenous constituent matter may compose as either a stuff mass or integral (count/number neutral uncountable) i-set. This is taken to explain why all fruit and vegetable terms in not only Welsh and Hejazi Arabic, but also Mandarin Chinese, have freely available stuff-readings, context permitting. The second method of achieving nominal flexibility, GRIND, is a COUNT → MASS shifting device which applies only to count terms ($\langle$**body**, **base**$\rangle$ or $\langle$***body**, **base**$\rangle$). It may be the case that not all languages have GRIND, but those which do seem to modulate its use with respect to context, last resort, or both. Due to the restrictions of GRIND, it is predicted that number neutral uncountable terms (which do not refer to fruits) should resit stuff-readings in and across languages (unless `ACCIDENTALLY HOMOPHONOUS ROOT CONCEPTS` applies).

Future work

The approach proposed in this work is not radical in the sense that Grimm (2012b) has already demonstrated that individuation and aggregation is crucial for the collective/singulative class, and also de Vries and Tsoulas (2021)'s Subpredicative Iceberg Semantics model already identifies **f:set* as locus of crosslinguistic variation for countability distinctions. What this work does is strips nominal number back to basics, identifying the key primitives that should underlie nominal number, and the features which are language specific. The power afforded by the Subpredicative Iceberg Semantics model proposed here is the architecture is solid: minimal semantic types and functions are proposed, *i.e.* the bare bones of the generative component and functions of number have been identified and developed. The future of the Subpredicative Iceberg Semantics model is seen to be easily modifiable over time as we learn more about diverse number systems. This can be done by adapting the core laid out in this thesis in a number of ways.

One way Subpredicative Iceberg Semantics can be developed further is by incorporating language specific nuances beyond the basic functions proposed, such as the derived collectives of the Slavic and Czech. The semantic core derived collectives may well be the same as the core of the Welsh plural (*i.e.* a PL_{NC}(P) derivation of a $\langle$***body**, ***base**$\rangle$ i-set), but there is much more to be said about the restrictiveness of these functions, and also the sense of clustering they impart, which clearly does not hold in Welsh. Similarly, the semantic core of specific classifiers in *e.g.* Mandarin

Chinese may well be the same as the semantic core for the singulative/general classifier function (*i.e.* a CL_{SING}(P) derivation of a $\langle$**body**, **base**$\rangle$ i-set), but there needs to be an explanation for the restrictiveness of these classifiers with respect to the types of nouns they are restricted to.

This work has focussed almost entirely on nominal terms for natural notional kinds. This is because the collective/singulative class covers these notional categories. Future work can, and should, incorporate other types of nouns into the system. English, for example, is known to have object mass terms which cover artificial superordinate collections (*e.g.* FURNITURE, POTTERY, CROCKERY). The incorporation of these terms into the theory of Subpredicative Iceberg Semantics may well involve further modification to the presuppositions associated with **f:set*.

There still remains features of this system that need to be identified, including the proper treatment of paucal semantics in *e.g.* Hejazi/Modern Standard Arabic. Further, the incorporation of the mechanism that derives exclusive plurals from inclusive plurals in downwards entailing and question contexts remains unsolved. Outside the primary focus of this thesis, there are still questions regarding how the subpredicative model accounts for abstract nouns, artificial collection nouns, and nouns with referents that don't exist in the real world, or exist only at the construal level.

Finally, a large part of this work assumes a specific model of the connection between construed ontological reality and its representation at the linguistic ontological level. In chapter 7, I touched upon the difference between substances and non-substances as being a clearly modal concept at some conceptual level. While I offered a sketch of an incorporation into Subpredicative Iceberg Semantics, the specifics of how this can be achieved remain to be seen.

Bibliography

Acquaviva, P. (2015). Singulatives. In Müller, P. O., Ohnheiser, I., Olsen, S., and Rainer, F., editors, *Word-Formation. An international handbook for the languages of Europe*, volume 2, pages 1171 – 1182. De Gruyter Mouton, Berlin, München, Boston.

Acuaviva, P. (2008). *Lexical Plurals: A Morphosemantic Approach.* Oxford University Press, Oxford.

Alexiadou, A. (2011). Plural mass nouns and the morpho-syntax of number. In Washburn, M. B., McKinney-Bock, K., Varis, E., Sawyer, A., and Tomaszewicz, B., editors, *Proceedings of West Coast Conference on Formal Linguistics 28*, pages 33 – 41.

Alexiadou, A. (2015). Deriving fake mass nouns. In Bui, T. and Özyıldız, D., editors, *Proceedings of the 45th North East Linguistics Society*, volume 45, pages 11–24.

Allan, K. (1980). Nouns and countability. *Language*, 56(3):541 – 567.

Alshboul, S., Asassfeh, S. M., and shaboul, Y. A. (2013). Default diminutives: Evidence from modern standard Arabic. *Journal of Language Teaching and Research*, 4(1):21–26.

Arensen, J. (1982). *Murle grammar. (Occasional Papers in the Study of Sudanese Languages, 2).* University of Juba.

Asmus, A. and Werner, E. (2015). Singulatives in modern Celtic and Slavic languages: Evidence from Welsh and Sorbian. In *Studia Celto-Slavica*, volume 7, pages 89 – 104.

Bach, E. (1986a). The algebra of events. *Linguistics and Philosophy*, 9(1):5–16.

Bach, E. (1986b). Natural language metaphysics. In Marcus, R. B., Dorn, G., and Weingartner, P., editors, *Logic, Methodology, and Philosophy of Science*, volume 7, pages 573–595.

Bale, A. and Barner, D. (2018). Quantity judgment and the mass-count distinction across languages: Advances, problems, and future directions for research. *Glossa: a journal of general linguistics*, 3(1).

Bale, A., Gagnon, M., and Khanjian, H. (2010). Cross-linguistic representations of numerals and number marking. In Li, N. and Lutz, D., editors, *Proceedings of Semantics and Linguistic Theory 20*, pages 582 – 598.

Bale, A. and Khanjian, H. (2014). Syntactic complexity and competition: The singular-plural distinction in Western Armenian. *Linguistic Inquiry*, 45(1):1 – 26.

Bale, A. C. and Barner, D. (2009). The interpretation of functional heads: Using comparatives to explore the mass/count distinction. *Journal of Semantics*, 26(3):217–252.

Barner, D., Inagaki, S., and Li, P. (2009). Language, thought, and real nouns. *Cognition*, 111(3):329 – 344.

Barner, D., Li, P., and Snedeker, J. (2010). Words as windows to thought: The case of object representation. *Current Directions in Psychological Science*, 19(3):195–200.

Barner, D. and Snedeker, J. (2005). Quantity judgments and individuation: Evidence that mass nouns count. *Cognition*, 97(1):41–66.

Beviláqua, K., Lima, S., and de Oliveira, R. P. (2016). Bare nouns in Brazilian Portuguese: An experimental study on grinding. In Rothstein, S. and Šķilters, J., editors, *The Baltic International Yearbook of Cognition, Logic and Communication*, volume XI. New Prairie Press.

Bittner, T. (2017). Vague distance predicates. *GeoInformatica*, 21(2):209–229.

Bliss, H. (2004). The semantics of the bare noun in Turkish. *Calgary Working Papers in Linguistics*, 25:1–65.

Bloomfield, L. (1933/1984). *Language*. University of Chicago Press.

Borer, H. (2005a). *In Name Only*, volume 1 of *Structuring Sense*. Oxford University Press, Oxford.

Borer, H. (2005b). *The Normal Course of Events*, volume 2 of *Structuring Sense*. Oxford University Press, Oxford.

Borer, H. and Ouwayda, S. (2010). Men and their apples: Dividing plural and agreement plural. Handout at Generative Linguistics in The Old World in Asia VIII.

Borer, H. and Ouwayda, S. (2021). Divide and counter. In Kiss, T., Pelletier, F. J., and Husić, H., editors, *Things and Stuff: The Semantics of the count/mass distinction*. Cambridge University Press.

Borsley, R. D., Tallerman, M., and Willis, D. W. E. (2007). *The Syntax of Welsh*. Cambridge University Press.

Bunt, H. C. (1985). *Mass terms and model-theoretic semantics*. Cambridge University Press, Cambridge.

Carey, S. and Xu, F. (2001). Infants' knowledge of objects: beyond object files and object tracking. *Cognition*, 80(1):179–213.

Casati, R. and Varzi, C. (1999). *Parts and places: The structures of spatial representation*. MIT Press, Cambridge, MA.

Champollion, L. (2010). *Parts of a whole: Distributivity as a bridge between aspect and measurement*. PhD thesis, University of Pennsylvania, Philadelphia, PA.

Champollion, L. and Krifka, M. (2016). Mereology. In Aloni, M. and Dekker, P., editors, *Cambridge handbook of formal semantics*. Cambridge University Press.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. University of California Press, Berkley.

Cheng, L. (2012). Counting and classifiers. In Massam, D., editor, *Count and Mass Across Langauges*, pages 199–219. Oxford University Press.

Cheng, L., Doetjes, J., and Sybesma, R. (2008). How universal is the Universal Grinder? *Linguistics in the Netherlands*, 25(1):50–62.

Cheng, L., Doetjes, J., Sybesma, R., and Zamparelli, R. (2012). On the interpretation of number and classifiers. *Italian Journal of Linguistics*, 24(2):175–194.

Cheng, L. and Sybesma, R. (1998). Yi-wan tag, yi-ge tang: Classifiers and massifiers. *Tsing Hua Journal of Chinese Studies*, XXVIII(3):385 – 412.

Cheng, L. and Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry*, 30(4):509–542.

Chierchia, G. (1998a). Plurality of mass nouns and the notion of "semantic parameter". In Rothstein, S., editor, *Events in Grammar.*, volume 70 of *Studies in Liguistics and Philosophy*, pages 53–103. Springer, Dordecht.

Chierchia, G. (1998b). Reference to kinds across language. *Natural Language Semantics*, 6(4):339–405.

Chierchia, G. (1999). Syntax-semantics interface. In Keil and Wilson, editors, *MIT Encyclopedia of Cognitive Science*. MIT Press.

Chierchia, G. (2010a). Language, thought, and reality after Chomsky. In Franck, J. and Bricmon, J., editors, *Chomsky Notebook*, Columbia Themes in Philosophy. Columbia University Press.

Chierchia, G. (2010b). Mass nouns, vagueness, and semantic variation. *Synthese*, 174:99–149.

Chierchia, G. (2015). How universal is the mass/count distinction? Three grammars of counting. In Li, Simpson, and Tsai, editors, *Chinese Syntax in a Cross-linguistic Perspective*. Oxford University Press.

Chierchia, G. (2017). Clouds and blood. More on vagueness and the mass/count distinction. *Synthese*, 194:2523–2538.

Chierchia, G. (2021). Mass vs. count: Where do we stand? outline of a theory of semantic variation. In Kiss, T., Pelletier, F. J., and Husić, H., editors, *Things and Stuff: The Semantics of the count/mass distinction*. Cambridge University Press.

Corbett, G. G. (1996). Minor number and the plurality split. *Italian Journal of Linguistics*, 8(1):101–122.

Corbett, G. G. (2000). *Number*. Cambridge University Press, Cambridge.

Cowell, M. W. (1964). *A Reference Grammar of Syrian Arabic (based on the dialect of Damascus)*. Georgetown University Press, Washington.

Cowper, E. and Hall, D. C. (2012). Aspects of individuation. In Massam, D., editor, *Count and Mass Across Languages*, pages 27–53. Oxford University Press.

Cuzzolin, P. (1998). Sull'origine del singolativo in Celtico, con particolare riferimento al medio Gallese. In *Archivio Glottologico Italiano*, volume 83, pages 121–149.

Dali, M. (2020). *Gender and Number in Tunisian Arabic: A Case of Allosemy*. PhD thesis, University of Ottawa.

Dali, M. and Mathieu, É. (2020). Broken plurals and (mis)matching of ϕ-features in Tunisian Arabic. *Brill's Journal of Afroasiatic Languages and Linguistics*, 12(2):164 – 203.

Dali, M. and Mathieu, É. (2021a). Singulative systems. In Hofherr, P. C. and Doetjes, J., editors, *The Oxford Handbook of Grammatical Number*. Oxford University Press.

Dali, M. and Mathieu, É. (2021b). *A Theory of Distributed Number*. John Benjamins Publishing Company.

de Belder, M. (2011). A morphosyntactic decomposition of countability in Germanic. *The Journal of Comparative Germanic Linguistics*, 14(3):173–202.

de Belder, M. (2013). Collective mass affixes: When derivation restricts functional structure. *Lingua*, 126:32–50.

de Vries, H. and Tsoulas, G. (2021). Overlap and countability in exoskeletal syntax: A best-of-both-worlds approach to the mass/count distinction. In Kiss, T., Pelletier, F. J., and Husić, H., editors, *Things and Stuff: The Semantics of the count/mass distinction*. Cambridge University Press.

de Vries, H. and Tsoulas, G. (2024). Portions and countability: A crosslinguistic investigation. *Natural Language & Linguistic Theory*, 42(1):383–435.

Deal, A. R. (2017). Countability distinctions and semantic variation. *Natural Language Semantics*, 25(2):125–171.

Dimmendaal, G. J. (2000). Number marking and noun categorization in Nilo-Saharan languages. *Anthropological Linguistics*, 42(2):214–261.

Doetjes, J. (2017). The count/mass distinction in grammar and cognition. *Annual Review of Linguistics*, 3:199–217.

Doetjes, J. (2021). Count-mass asymmetries: The importance of being count. In Kiss, T., Pelletier, F. J., and Husić, H., editors, *Things and Stuff: The Semantics of the count/mass distinction*. Cambridge University Press.

Erbach, K. (2021). *Object Mass Nouns: A Frame Based Analysis*. PhD thesis, Henrich Heine Universität Düsseldorf.

Erbach, K., Sutton, P. R., and Filip, H. (2019). Bare nouns and the Hungarian mass/count distinction. In Silva, A., Staton, S., Sutton, P., and Umbach, C., editors, *Language, Logic, and Computation*, pages 86–107. Springer Berlin Heidelberg, Berlin, Heidelberg.

Erbach, K., Sutton, P. R., Filip, H., and Byrdeck, K. (2017). Object mass nouns in Japanese. In Cremers, A., van Gessel, T., and Roelofsen, F., editors, *Proceedings of the 21st Amsterdam Colloquium*, pages 235 – 244. Institute for Logic, Language, and Computation at the University of Amsterdam.

Erschler, D. (2022). A dedicated nominal singular morpheme without singulative semantics. *Morphology*, 32(2):249–276.

Farkas, D. F. and de Swart, H. E. (2010). The semantics and pragmatics of plurals. *Semantics & Pragmatics*, 3:1–54.

Fassi Fehri, A. (2004). Nominal classes, reference and functional parameters, with particular reference to Arabic. *Linguistic Variation Yearbook*, 4(1):41–108.

Fassi Fehri, A. (2012). *Key Features and Parameters in Arabic Grammar*. John Benjamins Publishing Company.

Fassi Fehri, A. (2016). Semantic gender diversity and its architecture in the grammar of Arabic. *Brill's Journal of Afroasiatic Languages and Linguisticss*, 8(1):154–199.

Fassi Fehri, A. (2020). Number and gender convergence: The Arabic plurative. *Catalan Journal of Linguistics*, 19:75–126.

Gathercole, V. C. M., Thomas, E. M., and Evans, D. (2000). What's in a noun? Welsh-, English-, and Spanish- speaking children see it differently. *First Language*, 20(58):055–90.

Gil, D. (1996). Maltese "collective nouns": A typological perspective. *Italian Journal of Linguistics*, 8(1):53–87.

Gillon, B. S. (1992). Towards a common semantics for English count and mass nouns. *Linguistics and Philosophy*, 15:597–639.

Gleason, H. A. (1965). *Linguistics and English grammar*. Holt, Reinhart and Winston, New York.

Gleason, H. A. (1969). *An introduction to descriptive linguistics*. Henry Holt.

Gleitman, L. R. and Papafragou, A. (2005). Language and thought. In Holyoak, K. and Morrison, R., editors, *Cambridge handbook of thinking and reasoning*, pages 633–661. Cambridge University Press.

Goodman, N. (1951). *The structure of appearance*, volume 53. Harvard University Press.

Goodman, N. and Quine, W. V. O. (1947). Steps toward a constructive nominalism. *The Journal of Symbolic Logic*, 13(1):49–50.

Grätzer, G. (1998). *General Lattice Theory*. Birkhäuser Verlag, Base, second edition.

Greenberg, J. H. (1972). Numeral classifiers and substantival number: Problems in the genesis of a linguistic type. *Working Papers on Language Universals*, 9:1–39.

Grimm, S. (2010). Number and markedness: A view from Dagaare. In Prinzhorn, M., Schmitt, V., and Zobel, S., editors, *Proceedings of Sinn und Bedeutung 14*, pages 168 – 184.

Grimm, S. (2012a). Degrees of countability: A mereotopological approach to the mass/count distinction. In *Proceedings of Semantics and Linguistic Theory 22*, pages 584 – 603.

Grimm, S. (2012b). *Number and Individuation*. PhD thesis, Stanford University.

Grimm, S. (2018). Grammatical number and the scale of individuation. *Language*, 94(3):527 – 574.

Grimm, S. and Dočekal, M. (2021). Counting aggregates, groups and kinds: Countability from the perspective of a morphologically complex language. In Filip, H., editor, *Countability in Natural Language*, pages 85–120. Cambridge: Cambridge University Press.

Harbour, D. (2014). Paucity, abundance, and the theory of number. *Language*, 90(1):185–229.

Haspelmath, M. (2008). Ditransitive constructions: Towards a new role and reference grammar account? *Investigations of the Syntax–Semantics–Pragmatics Interface, 75-100 (2008)*.

Haspelmath, M. and Karjus, A. (2017). Explaining asymmetries in number marking: Singulatives, pluratives, and usage frequency. *Linguistics*, 55(6):1213–1235.

Hauser, M. and Spaulding, B. (2006). Wild rhesus monkeys generate causal inferences about possible and impossible physical transformations in the absence of experience. *Proceedings of the National Academy of Sciences*, 103(18):7181–7185.

Hauser, M. D. and Carey, S. (2003). Spontaneous representations of small numbers of objects by rhesus macaques: examinations of content and format. *Cogn Psychol*, 47(4):367–401.

Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598):1569–1579.

Hayward, D. (1984). *The Arbore language: A first investigation (including a vocabulary)*. Helmut Buske Verlag.

Haywood, J. A. and Nahmad, N. M. (1965). *A new Arabic grammar of the written language*. Lund Humphries, London, second edition.

Heim, I. (1991). Artikel und Definitheit. In von Stechow, A. and Wunderlich, D., editors, *Semantik: Ein internationales Handbuch de seitgenös-sichen Forschung*, pages 487–535. De Gruyter.

Heinz, S. (2009). Diminutive formations in modern Welsh and functions of their formatives. In Bloch-Trojnar, M., editor, *Perspectives on Celtic Languages*, Lublin Studies in Celtic Languages, pages 186–198. Wydawnictwo KUL.

Hemon, R. (1975). *A Historical Morphology and Syntax of Breton*, volume III of *Mediaeval and modern Breton series*. Dublin Institute for Advanced Studies.

Hemon, R. (2011). *Breton grammar. Translated, adapted and revised by Michael Everson*. Evertype, third edition.

Henderson, R. (2017). Swarms: Spatiotemporal grouping across domains. *Natural Language & Linguistic Theory*, 35(1):161–203.

Hnout, C., Laks, L., and Rothstein, S. (2021). Mass to count shifts in the Galilee dialect of Palestinian Arabic. In Kiss, T., Pelletier, F. J., and Husić, H., editors, *Things and Stuff: The Semantics of the count/mass distinction*. Cambridge University Press.

Iljic, R. (1994). Quantification in Mandarin Chinese: two markers of plurality. *Linguistics*, 32(1):91 – 116.

Imai, M. and Gentner, D. (1997). A cross-linguistic study of early word meaning: universal ontology and linguistic influence. *Cognition*, 62(2):169–200.

Inagaki, S. and Barner, D. (2009). Countability in absence of count syntax: Evidence from Japanese quantity judgments. *Studies in Language Sciences*, 8:111 – 125.

Ionin, T. and Matushansky, O. (2004). A singular plural. In Chand, V., Kelleher, A., Rodriguez, A., and Schmeiser, B., editors, *Proceedings of the 23rd West Coast Conference on Formal Linguistics*, volume 23, pages 399 – 412.

Ionin, T. and Matushansky, O. (2006). The composition of complex cardinals. *Journal of Semantics*, 23(4):315–360.

Ivlieva, N. (2013). *Scalar Implicatures and the Grammar of Plurality and Disjunction*. PhD thesis, Massachusetts institure of Technology.

Jackendoff, R. (1991). Parts and boundaries. *Cognition*, 41(1):9–45.

Jaradat, A. (2023a). Branching lexical plural into greater and paucal. *Cogent Arts & Humanities*, 10(1):2212462.

Jaradat, A. (2023b). When a dual marker acts as a paucal marker: The case of the dual -e:n in Northern Rural Jordanian Arabic. *Languages*, 8(3).

Jaradat, A. and Jarrah, M. (2022). The syntax of plurals of collective and mass nouns: Views from jordanian arabic. *Poznan Studies in Contemporary Linguistics*, 58(3):509–539.

Jespersen, O. (1924). *The Philosophy of Grammar*. Allen and Unwin, London.

Jones, M. and Thomas, A. (1977). *The Welsh Language: Studies in its Syntax and Semantics*. University of Wales Press for the Schools Council, Cardiff.

Joosten, F. (2003). Accounts of the count-mass distinction: A critical survey. *Lingvisticæ Investigationes*, 26(1):159–173.

Kagan, O. (2024). The Slavic suffix -in/-yn as partition shifter. *Natural Language Semantics*, 32(1):35–63.

Kagan, O. and Nurmio, S. (2024). Diminutive or singulative? the suffixes -in and -k in Russian. In Manova, S., Grestenberger, L., and Korecky-Kröll, K., editors, *Diminutives across Languages, Theoretical Frameworks and Linguistic Domains*, pages 65–88. De Gruyter Mouton, Berlin, Boston.

Kang, B.-M. (1994). Plurality and other semantic aspects of common nouns in Korean. *Journal of East Asian Linguistics*, 3(1):1–24.

Khrizman, K., Landman, F., Lima, S., Rothtein, S., and Schvarcz, B. R. (2015). Portion readings are count readings, not measure readings. In Brochhagen, T., Roelofsen, F., and Theiler, N., editors, *Proceedings of the 20th Amsterdam Colloquium*.

Kim, J. (2008). The parameterization of plural markings in classifier languages. *Kansas Working Papers in Linguistics*, 30.

King, G. (2003). *Modern Welsh: A Comprehensive Grammar*. Routledge, second edition.

Kiss, T., Pelletier, F. J., and Husić, H. (2021). Editorial introduction: Background to the count-mass distinction. In Kiss, T., Pelletier, F. J., and Husić, H., editors, *Things and Stuff: The Semantics of the count/mass distinction*. Cambridge University Press.

Kouneli, M. (2019). *The syntax of number and modification: an investiation of the Kipsigis DP*. PhD thesis, New York University.

Krifka, M. (1989). Nominal reference, temporal constitution and quantification in event semantics. In Bartsch, R., van Benthem, J., and van Emede Boas, P., editors, *Semantis and Contextual Expressions*, volume 11 of *Groningen-Amsterdamn Studies in Semantics*, pages 75–115. Foris Publications.

Krifka, M. (1995). Common nouns: A contrastive analysis of Chinese and Engish. In Carlson, G. and Pelletier, F. J., editors, *The Generic Book*, pages 398–411. University of Chicago Press.

Lahrouchi, M. and Ridouane, R. (2016). On diminutives in and plurals Moroccan Arabic. *Morphology*, 26(3-4):456–475.

Landman, F. (1989a). Groups, I. *Linguistics and Philosophy*, 12(5):559–605.

Landman, F. (1989b). Groups, II. *Linguistics and Philosophy*, 12(5).

Landman, F. (2011). Count nouns, mass nouns, neat nouns, mess nouns. *Baltic International Yearbook of Cognition, Logic and Communication*, 6.

Landman, F. (2016). Iceberg semantics for count nouns and mass nouns: Classifiers, measures and portions. *Baltic International Yearbook of Cognition, Logic and Communication*, 11.

Landman, F. (2020). *Iceberg Semantics for Mass Nouns and Count Nouns*, volume 105 of *Studies in Linguistics and Philosophy*. Springer International Publishing.

Landman, F. (2021). Iceberg semantics for count nouns and mass nouns: How mass counts. In Filip, H., editor, *Countability in Natural Language*. Cambridge University Press.

Lea, S. E. G., Slater, A. M., and Ryan, C. M. E. (1996). Perception of object unity in chicks: A comparison with the human infant. *Infant Behavior & Development*, 19(4):501–504.

Leonard, H. S. and Goodman, N. (1940). The calculus of individuals and its uses. *The Journal of Symbolic Logic*, 5(2):45–55.

Leśniewski, S. (1916). Podstawy ogólnej teoryi mnogości. i. *Prace Polskiego Koła Naukowego w Moskwie*, 2:5–42.

Li, P., Dunham, Y., and Carey, S. (2009). Of substance: The nature of language effects on entity construal. *Cognitive Psychology*, 58(4):487–524.

Li, Y.-H. A. (1999). Plurality in a classifier language. *Journal of East Asian Linguistics*, 8(1):75–99.

Lima, S. (2010). About the mass-count distinction in Yudja: a description. In Rogers, B. and Szakay, A., editors, *Proceedings of the 15th Workshop on Structure and Constituency in Languages of the Americas*.

Lima, S. (2014). All notional mass Nouns are count nouns in Yudja. *Proceedings of Semantics and Linguistic Theory 24*, 24.

Link, G. (1983). The logical analysis of plural and mass terms. In Bäuerle, R., Schwarze, C., and von Stechow, A., editors, *Meaning, Use, and Interpretation of Language*. Walter de Gruter, Berlin.

Longobardi, G. (1996). The Sytax of N-raising: A Minimalist Theory. In *OTS Working Papers, Research Institute for Language and Speech,*. University of Utrecht.

Lucy, J. A. (1992). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge University Press.

Lucy, J. A. and Gaskins, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In Bowerman, M. and Levinson, S. C., editors, *Language acquisition and conceptual development*, Language, culture and cognition. Cambridge University Press, Cambridge.

Lucy, J. A. and Gaskins, S. (2003). Interaction of language type and referent type in the development of nonverbal classification preferences. In Gentner, D. and Goldin-Meadow, S., editors, *Language in mind: Advances in the issues of language and thought*, pages 465–492. MIT Press.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, 14(1):57–77.

Martí, L. (2017). Inclusive plurals and the theory of number. Pre-publication of Marti 2020.

Martí, L. (2020). Inclusive plurals and the theory of number. *Linguistic Inquiry*, 51:37–74.

Mathieu, É. (2012a). Flavours of division. *Linguistic Inquiry*, 43(4):650–679.

Mathieu, É. (2012b). On the mass/count distinction in Ojibwe. In Massam, D., editor, *Count and Mass Across Languages*. Oxford University Press.

Mathieu, É. (2013). On the plural of the singulative. *McGill Working Papers in Linguistics*, 23(1).

Mathieu, É. (2014). Many a plural. In Aguilar-Guevara, A., Le Bruyn, B., and Zwarts, J., editors, *Weak Referentiality*, volume 219, pages 157–182. John Benjamins Publishing Company.

Mathieu, É. (2018). Humans, gods, and demons. In Mathieu, É., dali, M., and Zareikar, G., editors, *Gender and Noun Classification*, pages 1–14. Oxford University Press.

Mathieu, É. and Dali, M. (2021). The semantics of distributed number. In Kiss, T., Pelletier, F. J., and Husic, H., editors, *Things and Stuff: The Semantics of the count/mass distinction*. Cambridge University Press.

Matthews, P. H. (2014). *The concise Oxford dictionary of linguistics*. Oxford University Press, Oxford, third edition.

Mayr, C. (2015). Plural definite NPs presuppose multiplicity via embedded exhaustification. In D'Antonio, S., moroney, M., and Little, C. R., editors, *Proceedings of Semantics and Linguistic Theory 25*, volume 25, pages 204 – 224.

McCarthy, J. J. and Prince, A. S. (1990). Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language & Linguistic Theory*, 8(2):209–283.

McCawley, J. D. (1975). Lexicography and the count-mass distinction. In *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, pages 314–321.

Middleton, E., Wisniewski, E., Trindel, K., and Imai, M. (2004). Separating the chaff from the oats: Evidence for a conceptual distinction between count noun and mass noun aggregates. *Journal of Memory and Language*, 50:371–394.

Mittendorf, I. and Sadler, L. (2005). Numerals, nouns and number in Welsh NPs. In Butt, M. and King, T. H., editors, *Proceesings of the LFG'05 Conference*, Standford, California. CSLI Publications.

Moltmann, F. (1998). Part structures, integrity, and the mass-count distinction. *Synthese*, 116(1):75–111.

Moltmann, F. (2020). Introduction. In Moltmann, F., editor, *Mass and Count in Linguistics, Philosophy, and Cognitive Science*, pages 2–12. Benjamins.

Moltmann, F. (2021). Levels of ontology and natural language: The case of the ontology of parts and wholes. In Miller, J., editor, *The Language of Ontology*. Oxford University Press.

Nurmio, S. (2017). Collective nouns in Welsh: A noun category or a plural allomorph? *Transactions of the Philological Society*, 115:1:58–78.

Nurmio, S. (2020). *Grammatical Number in Welsh: Diachrony and Typology*, volume 117 of *Transactions of the Philological Society*. John Wiley & Sons.

Nurmio, S. (2023). Towards a typology of singulatives: Definition and overview of markers. In Arbes, D., editor, *Number Categories: Dynamics, Contact, Typology*, pages 155–182. De Gruyter Mouton.

Ojeda, A. E. (1992). The semantics of number in Arabic. In *Proceedings of Semantics and Linguistic Theory 2*.

Ouwayda, S. (2014). *Where Number Lies: Plural Marking, numerals, and the collective-distributive distinction*. PhD thesis, University of Southern California.

Palmer, F. R. (1990). The semantics of grammar. *Journal of Linguistics*, 26(1):223–233.

Park, S.-Y. (2008). Plural marking in classifier languages: a case study of the so-called plural marking -tul in Korean. In *Toronto Working Papers in Linguistics*, volume 280.

Partee, B. H., ter Muelen, A. G. B., and Hall, R. E. (1990). *Mathematical Methods in Linguistics*. Dordrecht, Boston.

Pelletier, F. J. (1975). Non-singular reference: Some preliminaries. *Philosophia*, 5:451–546.

Pelletier, F. J. (1979). Non-singular reference: Some preliminaries. In Pelletier, F. J., editor, *Mass Terms: Some Philosophical Problems*, pages 1–14. Springer Netherlands, Dordrecht.

Pelletier, F. J. (2011). Descriptive metaphysics, natural language metaphysics, Sapir-Whorf, and all that stuff: Evidence from the mass-count distinction. In Rothstein, S. and Šķilters, J., editors, *The Baltic International Yearbook of Cognition, Logic and Communication*, volume 6. New Prairie Press.

Pelletier, F. J. (2012). Lexical nouns are both +mass and +count, but they are neither +mass nor +count. *Education and Research Archive, University of Alberta*.

Press, I. (2004). *Standard Breton*. Languages of the world:/Materials. Lincom Europa.

Quine, W. V. O. (1960). *Word and Object*. MIT Press, Cambridge MA.

Quine, W. V. O. (1969). *Ontological relativity and other essays*. The John Dewey Essays in Philosophy. Columbria university Press.

Reh, M. (1985). *Die Krongo-Sprache (nìino mó-di): Beschreibung, Texte, Wörterverzeichnis*. Reimer.

Renans, A., Sağ, Y., Ketrez, F. N., Tieu, L., Tsoulas, G., Folli, R., de Vries, H., and Romoli, J. (2020). Plurality and crosslinguistic variation: an experimental investigation of the Turkish plural. *Natural Language Semantics*, 28(4):307–342.

Rescher, N. (1955). Axioms for the part relation. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 6(1):8–11.

Roberts, S. and Gathercole, V. (2012). Talking of objects: How different are Welsh and English nouns? *Journal of Celtic Linguistics*, 14(1):67–85.

Rosiak, K. (2009). The use of diminutives by speakers of Welsh. In Bloch-Trojnar, M., editor, *Perspectives on Celtic Languages*, pages 237–247. Wydawnictwo KUL.

Rosiak, K. (2013). From synthetic to analytic? - the changing use of diminutive expressions in Welsh. In Ahiqvist, A. and O'Neill, P., editors, *Celts and their Cultures at home and abroad. A Festschrift for Malcolm Broun.*, pages 289–309. The Celtic Studies Foundation, the University of Sydney.

Rothstein, S. (2009). Individuating and measure readings of classifier constructions: Evidence from Modern Hebrew. *Brill's Journal of Afroasiatic Languages and Linguistics*, 1(1):106–145.

Rothstein, S. (2010). Counting and the mass/count distinction. *Journal of Semantics*, 27(3):343–397.

Rothstein, S. (2011). Counting, measuring and the semantics of classifiers. *Baltic International Yearbook of Cognition, Logic and Communication*, 6.

Rothstein, S. (2017). *Semantics for Counting and Measuring*. Key Topics in Semantics and Pragmatics. Cambridge University Press.

Rothstein, S. (2021). Counting, plurality, and portions. In Kiss, T., Pelletier, F. J., and Husić, H., editors, *Things and Stuff: The Semantics of the count/mass distinction*. Cambridge University Press.

Rullman, H. and You, A. (2006). General number and the semantics and pragmatics of indefinite bare nouns in Mandarin Chinese. In von Heusinger, K. and Turner, K., editors, *Where semantics meets pragmatics*, pages 175–196. Elsevier.

Sağ, Y. (2022). Bare singulars and singularity in Turkish. *Linguistics and Philosophy*, 45(4):741–793.

Samuelson, L. K. and Smith, L. B. (1999). Early noun vocabularies: do ontology, category structure and syntax correspond? *Cognition*, 73(1):1–33.

Santos, L. R. (2004). 'Core knowledges': a dissociation between spatiotemporal knowledge and contact-mechanics in a non-human primate? *Dev Sci*, 7(2):167–174.

Sapir, E. (1929). The status of linguistics as a science. *Language*, 5.

Sapir, E. (1944). Grading, a study in semantics. *Philosophy of Science*, 11(2):93–116.

Sauerland, U., Anderssen, J., and Yatushiro, K. (2005). The plural is semantically unmarked. In Kepser, S. and Reis, M., editors, *Linguistic Evidence: Empirical,: Theoretical and Computational Perspectives*. Mouton de Gruyter.

Scrontas, G. (2014). *The Semantics of Measurement*. PhD thesis, Harvard University.

Scwarzchild, R. (2011). Stubborn distributivity, multiparticipant nouns and the count/mass distinction. In Suzi Lima, K. M. and Smith, B. W., editors, *Proceedings of NELS 39*.

Seifart, F. (2005). *The structure and use of shape-based noun classes in Miraña (north west Amazon).* PhD thesis, Radboud Universiteit Nijmegen.

Sharvey, R. (1978). Maybe English has no count nouns: notes on Chinese semantics. *Studies in Language*, 2(3):345–265.

Simons, P. (1991). Part/whole II: Mereology since 1900. In Burkhardt, H. and Smith, B., editors, *Handbook of Metaphysics and Ontology*, pages 672–675. Philosophia, Munich.

Simpson, A. (2008). Classifiers and DP structure in Southeast Asia. In Cinque, G. and Kayne, R., editors, *The Oxford Handbook of Comparative Syntax*. Oxford University Press.

Smith, B. (1996). Mereotopology: A theory of parts and boundaries. *Data and Knowledge Engineering*, 3:287–303.

Smith-Stark, T. C. (1974). The plurality split. In La Galy, M. W., Fox, R. A., and Bruck, A., editors, *Papers from the Tenth Regional Meeting*. Chicago Linguistic Society.

Soja, N. N., Carey, S., and Spelke, E. S. (1991). Ontological categories guide young children's inductions of word meaning: Object terms and substance terms. *Cognition*, 38(2):179–211.

Spector, B. (2007). Aspects of the pragmatics of plural morphology: On higher-order implicatures. In Sauerland, U. and Stateva, P., editors, *Presupposition and Implicature in Compositional Semantics*, pages 243–281. Palgrave Macmillan UK, London.

Spelke, E. S. and Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, 10(1):89–96.

Stolz, T. (2001). Singulative-collective: Natural morphology and stable classes in Welsh number inflexion on nouns. *STUF - Language Typology and Universals*, 54(1):52–76.

Strawson, P. (1959). *Individuals*. Methuen, London.

Stump, G. (2005). Word-formation and inflectional morphology. In Štekauer, P. and Lieber, R., editors, *Handbook of word-formation*, pages 49–71. Springer, Dordrecht.

Sutton, P. R. and Filip, H. (2016a). Countability: Individuation and context. In Foret, A., Morrill, G., Muskens, R., Osswald, R., and Pogodalla, S., editors, *Formal Grammar*, pages 290–305, Berlin, Heidelberg. Springer Berlin Heidelberg.

Sutton, P. R. and Filip, H. (2016b). Counting in context: count/mass variation and restrictions on coercion in collective artifact nouns. *Proceedings of SALT 26*, page 350.

Sutton, P. R. and Filip, H. (2016c). Mass/count variation: A mereological, two-dimensional semantics. *Baltic International Yearbook of Cognition, Logic and Communication*, 11.

Sutton, P. R. and Filip, H. (2016d). Vagueness, overlap, and countability. In Bade, N., Berezovskaya, P., and Schöller, A., editors, *Proceedings of Sinn und Bedeutung 20*, pages 730–747.

Sutton, P. R. and Filip, H. (2017). Restrictions on subkind coercion in object mass nouns. In Truswell, R., Cummons, C., Haycock, C., Rabern, B., and Rohde, H., editors, *Proceedings of Sinn und Bedeutung 21*, pages 1195–1214.

Tarski, A. (1935). Zur Grundlegung der Booleschen Algebra I. In *Fundamenta mathematica*, volume 35. Oxford University, 1st ed edition.

Tiersma, P. M. (1982). Local and general markedness. *Language*, 58(4):832–849.

Tsoulas, G. (2009). On the grammar of number and mass terms in Greek. In *Proceedings of the 2007 Workshop in Greek Syntax and Semantics at MIT*.

Unseth, P. (1988). Majang nominal plurals, with comparative notes. *Studies in African Linguistics*, 19(1):75–91.

Valenza, E., Leo, I., Gava, L., and Simion, F. (2006). Perceptual completion in newborn human infants. *Child Dev*, 77(6):1810–1821.

Varzi, A. (2019). Mereology. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Varzi, A. C. (1998). Basic problems of mereotopology. In Guarino, N., editor, *Formal Ontology in Information Systems*, pages 29–38. IOS Press.

Varzi, A. C. (2007). Spatial reasoning and ontology: Parts, wholes, and locations. In Aiello, M., Pratt-Hartmann, I., and Van Benthem, J., editors, *Handbook of Spatial Logics*, pages 945–1038. Springer Netherlands, Dordrecht.

Wągiel, M. (2021a). Slavic derived collective nouns as spatial and social clusters. In Dočekal, M. and Wągiel, M., editors, *Formal approaches to number in Slavic and beyond*. Language Science Press.

Wągiel, M. (2021b). *Subatomic quantification*. Number 6 in Open Slavic Linguistics. Language Science Press, Berlin.

Wągiel, M. and Shlikhutka, N. (2022). A mereotopological account of Ukrainian singulatives. A talk presented at the 15th European Conference on Formal Description of Slavic Languages (FDSL 15).

Wągiel, M. and Shlikhutka, N. (2023a). Clustering and declustering things: The meaning of collective and singulative morphology in Ukrainian. In Kim, J., Öney, B., Zhang, Y., and Zhao, F. L., editors, *Proceedings of Semantics and Linguistic Theory 33*.

Wągiel, M. and Shlikhutka, N. (2023b). Parts of clusters and Ukrainian singulatives. A talk presented at the Deutsche Gesellschaft für Sprachwissenschaft 45 conference (DGfS 45).

Wągiel, M. and Shlikhutka, N. (2023c). Pluralities, clusters and singulatives: A view from Ukrainian. A talk presented at GLOW 46.

Ware, R. X. (1979). Some bits and pieces. In Pelletier, F. J., editor, *Mass Terms: Some Philosophical Problems*, pages 15–29. Springer Dordecht, Dordrecht.

Wehr, H. (1961). *A Dictionary of Modern Written Arabic*. Harrassowitz Verlag.

Whitehead, A. N. (1920). *The concept of nature*. Cambridge University Press.

Whitehead, A. N. (1929). *Process and reality: An Essay in Cosmology*. Macmillan.

Whorf, B. L. (1940a). Linguistics as an exact science. *Technology Review*, 43.

Whorf, B. L. (1940b). Science and linguistics. *Technology Review*, 42.

Whorf, B. L. (1956). *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. The MIT Press.

Wierzbicka, A. (1988). *The Semantics of Grammar*. Benjamins, Amsterdam; Phiadelphia.

Wilhelm, A. (2008). Bare nouns and number in Dëne Sųłiné. *Natural Language Semantics*, 16(1):39–68.

Wiltschko, M. (2008). The syntax of non-inflectional plural marking. *Natural Language & Linguistic Theory*, 26(3):639–694.

Wright, W. (1896). *A grammar of the Arabic Language*, volume 1. Cambridge England : University Press.

Wu, Y.-C. Y. (2019). Plural classifier xie and grammatical number in Mandarin Chinese. In *Berkeley Papers in Formal Linguistics*, volume 2.

Xu, F., Carey, S., and Welch, J. (1999). Infants' ability to use object kind information for object individuation. *Cognition*, 70(2):137–166.

Yang, R. (2001). *Common nouns, classifiers, and quantification in Chinese*. PhD thesis, Rutgers University.

Yi, B.-U. (2009). Chinese classifers and count nouns. *Journal of Cognitive Science*, 10(2):209–225.

Zabbal, Y. (2002). The semantics of number in the Arabic number phrase. Master's thesis, University of Alberta.

Zhang, N. N. (2013). *Classifier Structures in Mandarin Chinese*, volume 263 of *Trends in Linguistics. Studies and Monographs.* De Gruyter Mouton, Berlin, Boston.

Zweig, E. (2009). Number-neutral bare plurals and the multiplicity implicature. *Linguistics and Philosophy*, 32(4):353–407.

Zwicky, A. (2001). Counting chad. Handout from Standforf Semantics Fest.

Appendix A

Collective/Singulative and Singular/Plural Word Lists

The following are non-exhaustive word lists for natural kinds in Welsh and Arabic sorted into the countability classes singular/plural and collective/singulative. The notional categories included cover *granular materials, fruits, vegetables, insects, crustaceans, animals, trees, and flowers*. I am using an intuitive classification of terms within these categories, rather than scientific classifications. As such, I do not distinguish between *insects, molluscs* and *arachnids*, nor do I distinguish *e.g.* a *banana* as a *berry*.

The lists are sorted by category, then nominal class, and then size of referents. I have applied an intuitive approach to size, rather than an absolute definition. Further information may be given regarding certain terms in footnotes. I also employ the following key:

† terms that are listed in at least one of the consulted dictionaries, but informants have indicated that the term is awkward or rarely used.

♥ terms where (some) speakers attested a strong preference for the collective form over the singulative form in all contexts.

♣ (Arabic only) The term is collective, but some speakers preferred to use the classifier HABA حبة rather than the singulative to individuate.

♦ terms have been borrowed

♠ terms which are compounds

These lists were compiled in consultation with native speakers of Welsh and native speakers of Arabic of varying dialects:

- Welsh Speakers
  - 2 x Northern Wales Welsh speaker
  - 1 x Southern Wales Welsh speaker (Professional English/Welsh translator)
- Arabic Speakers
  - 1 x Libyan Arabic Native speaker
  - 1 x Egyptian Arabic speaker (Professional English/MSA translator)
  - 1 x Hejazi Arabic (Saudi Arabia) speaker
  - 1 x Najdi Arabic (Saudi Arabia) speaker

Though care has been taken to ensure the accuracy of these lists, there is speaker and dialect variation, especially with regard to the Arabic entries. Speakers did not agree on all the terms listed. I have indicated where my consultants disagreed on certain terms, and added explanations in the footnotes. As there is speaker variation, these lists are not to be taken as definitive of the collective/singulative class for any given speaker/dialect of Welsh or Arabic, but rather indicative of the spread of collective/singulative versus singular/plural distribution.

The terms in these lists are, to the best of my knowledge, either standard varieties (Standard Welsh, Modern Standard Arabic) or commonly used forms across most dialects. Some dialectal variants may be present. Where a term is known to be dialectal variant, this is indicated in the lists. As such, some cases may have multiple terms for the same entities. As the terms in these lists have been curated to show standard/commonly used variants, the Arabic transliterations do not represent any specific accent or pronunciation.

With respect to Arabic singular/plural terms, not every plural is given. Singulars are given in masculine form.

Finally, a term being listed as being part of the collective/singulative class does not rule out that the terms have alternate plural forms that may be used. For example, SHAJAR → SHAJARAH shows the collective/singulative contrast, but the broken plural of the singulative A'SHJAAR is commonly used.

All terms have been cross-checked with entries in the following dictionaries:

- Welsh dictionaries consulted:
  - Geiriadur yr Academi (Bruce Griffiths and Dafydd Glyn Jones, University of Wales Press)[1]
  - Geiriadur Prifysgol Cymru (Prifysgol Cymru/The University of Wales)[2]
  - Geiriadur Ar-lein Cymraeg-Saesneg/ Saesneg-Cymraeg (Prifysgol Cymru, Y Drindod Dewi Sant/The University of Wales, Trinity Saint David)[3]
- Arabic dictionaries consulted:
  - *The Living Arabic Project*[4], an online community-sourced dictionary of Classical Arabic and Arabic dialects.
  - The Hans Wehr Dictionary of Modern Written Arabic, Edited by J.M. Cowan

[1]Available online: https://geiriaduracademi.org
[2]Available online: https://www.geiriadur.ac.uk
[3]Available online: https://geiriadur.uwtsd.ac.uk
[4]Available online: https://www.livingarabic.com

A.1 Welsh word list

Abbreviations: (N) = Northern Welsh, (S) = Southern Welsh, (Lit) = literary use

Granular Materials

Collective/Singulative System

llwch → llychyn	dust → a speck of dust
tywod → tywodyn	sand → a grain of sand
graean → greyenyn	gravel → a grain of gravel

Faux singulatives

cerigyn → cerigos †	pebble → pebbles[5]

Food Items

Grains, seeds & pulses

Collective/Singulative System

barlys → barlysyn (S)	barley → a grain of barley
gwenith → gwenithen	wheat → a grain of wheat
dincod → dincodyn	seeds → a seed[6]
ŷd → ydyn	corn → a piece of corn
ceirch → ceirchen	oats → an oat
corbys → corbysen	lentils/pulses → a lentil
ffacbys → ffacbysen ♠	lentils/chickpeas → a lentil
pys → pysen ♦	peas → a pea
ffa → ffeuen	beans → a bean
cidnabêns → cidnabensen (S) ♦	kidney beans → a kidney bean

[5] CARREG GRON → CERRIG CRYNION is more common

[6] The pair HAD(AU)/HEDYN may be used for seed/pip. Note HAD is collective while HADAU is plural.

Nuts

Collective/Singulative System

cnau → cneuen	nuts → a nut
cnau almon → cneuen almon ♠	almonds → an almond
cnau Ffrengig → cneuen Ffrengig ♠	walnuts → a walnut
cnau pistasio → cneuen bistasio ♠	pistachios → a pistachio
cnau pecan → cneuen becan ♠	pecans → a pecan
pysgnau → psygneuen ♠	peanuts → a peanut

Singular/plural system

almon → almonau ♦	almond → almonds
pecan → pecanau ♦	pecan → pecans

Berries and aggregate fruit

Collective/Singulative System

mwyar → mwyaren	berries→ berry
aeron → aeronen	berries → a berry
llugaeron → llugaeronen	cranberries → a cranberry
cyrains → cyransen	currents → a current
cyrains duon → cyransen ddu ♠	blackcurrants → a blackcurrant
cyrains cochion → cyransen goch ♠	redcurrants → a redcurrant
mafon→ mafonen	raspberries → raspberry[7]
mwyar duon → mwyaren ddu	blackberries → blackberry
grawnwin→ grawnwinen	grapes → a grape
grêps → grepsen ♦	grapes → a grape
mefus → mefusen (N)	strawberries → a strawberry[8]

Faux singulatives

resinen → resins ♦	raisin → raisins

[7] The pair AFAN → AFANEN is lesser used collective/singulative pair to refer to *raspberries*

[8] The collective/singulative pair SYFI → SYFÏEN are lesser used terms to refer to *strawberries*

Root Vegetables

Collective/Singulative System

rhuddygl → rhuddyglen	radishes → a radish
betys → betysen	beetroots → a beetroot
bitrwd → bitrwden	beetroots → a beetroot
moron → moronen	carrots → a carrot
caretsh → caretsien (N) ♦	carrots → a carrot
pannas → panasen	parsnips → a parsnip
erfin → erfinen (S)	turnips → a turnip
maip → meipen (N)	turnips → a turnip
sibols → sibolsyn (N)	shallots → a shallot
winwns → winwnsyn (S) ♠	onions → an onion
winwyn → winionyn (Lit) ♠	onions → an onion
cennin → cenhinen	leeks → a leek

Faux singulatives

nionyn → nionod (N) ♠	onion → onions
taten → tatws ♠	potato → potatoes

Cruciferous Vegetables

Collective/Singulative System

ysgewyll → ysgewyllen	sprouts → a sprout
letys → letysen ♦	lettuces → a lettuce
bresych → bresychen (Lit) †	cabbages → a cabbage[9]
blodfresych → blodfresychen	cauliflowers/broccolies → cauliflower/broccoli
cabaets → cabatsien ♦	cabbages → a cabbage

Singular/plural system

[9]Speakers questioned whether this was literary.

colifflŵar → colifflŵars ♦	cauliflower → cauliflowers
brocoli → brocolis ♦	broccoli → broccolis

Stone Fruit

Collective/Singulative System

ceirios→ ceiriosen	cherries→ a cherry
datys → datysen (Lit) ♦	dates → a date
dêts → deten ♦	dates → a date
bricyll → bricyllen	apricots → an apricot
eirin → eirinen (N)	plums → a plum
plyms → plymsen ♦	plums → a plum
eirin wlanog → eirinen wlanog ♠	peaches → a peach[10]

Singular/plural system

olif → olifau ♦	olive → olives
nectarîn → nectarinau ♦	nectarine → nectarines
afocado → afocados ♦	avocado → avocados
mango → mangoau ♦	mango → mangoes

Other Fruits and Vegetables

Collective/Singulative System

madarch → madarchen	mushrooms → a mushroom
ffigys → ffigysen ♦	figs → a fig
gellyg → gellygen	pears → a pear
pêrs → peren (N) ♦	pears → a pear

Singular/plural system

ciwi → ciwïod ♦	kiwi → a kiwi

[10] EIRIN/PEACH (collective) + WLANOG/WOOLLY.

afal → afalau	apple → apples
oren → orennau	orange → oranges
lemon → lemonau ♦	lemon → lemons
tomato → tomatos ♦	tomato → tomatoes
pupur → puprau ♦	peppers → a pepper
banana → bananas ♦	banana → bananas
pomgranad → pomgranadau ♦	pomegranates → a pomegranate
grawnffrwyth → grawnffrwythau ♦	grapefruit → grapefruits[11]
pîn-afal → pîn-afalau	pineapple → pineapples
ciwcymbr → ciwcymbrau ♦	cucumber → cucumbers
corbwmpen → corbwmpenni ♠	courgette → courgettes
melon → melonau♦	melons → a melon
dyfrfelon → dyfrfelonau ♠	watermelons → a watermelon[12]
maro → maros ♦	marrow → marrows
pympen→ pwmpenni ♦	pumpkin → pumpkins

Animals and Insects

Wingless Insects (including Arachnids and Molluscs)

Collective/Singulative System:

gwiddon → gwiddonyn	mites → a mite
chwain → chwannen	fleas → a flea
lleu → lleuen	louse → a lice
mogrug → morgrugyn	ants → an ant
myrion → myrionen (S)	ants → an ant
cynrhon → cynrhonyn	maggots → a maggot
lindys → lindysyn	caterpillars → a caterpillar
abwyd → abwydyn †	worms → a worm[13]

[11] GRAWN/GRAPES + FFRWTH/FRUIT (singular/plural).
[12] DYFR/WATER + FELON/MELON (singular/plural).
[13] The faux singulative pair is reported as more common.

malwod → malwoden (S) snails → a snail

pryfed cop → pryf copyn (N) ♠ spider → a spider[14]

Faux singulatives:

trogen → trogod tick → ticks

mwydyn → mwydod worm → worms

malwen → malwod (N) snail → snails

gwlithen → gwlithod slug → a slug

Singular/Plural System:

miltroed → miltroediaid millipede → millipedes

corryn → corynod spider → spiders

sgorpion → sgorpionau ♦ scorpion → scorpions

Winged Insects

Collective/Singulative System:

gwybed → gwybedyn gnats → a gnat

cylion → cylionen (S.W) flies → a fly

clêr → cleren (S) flies → a fly

gwenyn → gwenynen bees → a bee

cacwn → cacynen wasps → a wasp

gwenyn meirch → gwenynen feirch wasps → wasp

Faux singulatives:

sioncyn gwair→ sioncod gwair grasshopper → grasshoppers

chwilen ddu→ chwilod duon cockroach → cockroaches

chwilen → chwilod beetle → beetles

Singular/Plural System:

mosgito → mosgitos ♦ mosquito → mosquitos

[14] PRYFED/INSECT + COP/SPIDER. Note, some speakers drop PRYFED, making COP/COPYN their collective/singulative contrast.

locust → locustiaid ♦	locust → locusts
glöyn byw → glöynnod byw	butterlfy → butterflies[15]
pili-pala → pili-palod (S)	butterfly → butterflies

Other Insects

Singular/Plural System:

pryf → pryfed	insect → insects

Crustaceans

Collective/Singulative System:

berdys → berdysen	shrimps → a shrimp
wystrys → wystrysen	oysters → oyster

Singular/plural System:

corgimwch → corgimychiaid	prawn → prawns
cranc → crancod	crab → crabs
cimwch → cimychiaid	lobster → lobsters

Fish:

Collective/Singulative System:

pysgod → pysgodyn	fishes → a fish

Faux singulatives:

eog(yn) → eogiaid	salmon → salmons[16]
lleden → lledod	flounder → flounders

[15] Literally *living coal*: GLOYN/COAL.SING + BYW/LIVE.

[16] EOG is more accepted as the singular than EOGYN.

Singular/Plural System:

brwyniad → brwyniaid	anchovy → anchovies
penhwyad → penhwyaid	pike → pikes
penfras → penfreision	cod → cods
draenog → draenogiaid	perch → perches
cerpyn → cerpynnod	carp → carps
llysywen → llysywennod	eel → eels

Amphibians and Reptiles

Collective/Singulative System:

nadredd → nadredden (S.E) ✝	snakes → a snake[17]

Singular/Plural System:

amffibiad → amffibiaid	amphibian → amphibians
ymlusgiad → ymlusgiaid	reptile → reptiles
neider → nadredd	snake → snakes
penbwl → penbyliaid	tadpole → tadpoles
llyffant → llyffaint	frog/toad → frogs/toads[18]
madfall dŵr → madfallod dŵr	newt → newts
salamandr → salamandrau ♦	salamander → salamanders
crwban → crwbanod	turtle → turtles
igwana → igwanaod ♦	iquana → iguanas

Birds:

Collective/Singulative System:

adar → aderyn	birds → a bird
petris → petrisen	partridges → partridge

[17] Singular/plural is more commonly accepted.
[18] BROGA → BROGAOD is more common in S.W. Welsh.

adar duon → aderyn du	blackbirds → a blackbird
adar y to → aderyn y to	sparrows → a sparrow
hwyaid → hwyaden	ducks → duck
corhwyaid → corhwyaden	teals → a teal
pïod → pioden	magpies → a magpie

Singular/Plural System:

coblyn → coblynnod	woodpecker → woodpeckers
cigfran → cigfrain	raven → ravens
colomen → colomennod	pigeon → pigeons
cyw→ cywion	chicken → chickens
ceiliog → ceiliogod	rooster → roosters
gŵydd → gwyddau	goose → geese
tylluan → tylluanod	owl → an owl
alarch → elyrch	a swan → swans
fwltur → fwlturiaid	vulture → vultures
eryr → eryrod	eagle → eagles

Mammals:

Collective/Singulative System:

llygod → llygoden	mice → a mouse
llygod mawr → llygoden fawr	rats → a rat[19]
llygod y gwair → llygoden y gwair	voles → a vole
llygod bochog → llygoden fochog	hamsters → a hamster
moch cwta → mochyn cwta	ginuea pigs→ a ginea pig[20]
moch daear → mochyn daear	badgers → a badger[21]
moch → mochyn	pig → pigs

[19] LLYGOD/MOUSE + MAWR/BIG.
[20] MOCH/PIG + GWTA/SHORT.
[21] MOCH/PIG + DAEAR/EARTH.

Faux singulatives:

ystlum(yn) → ystlumod	bat → bats
cwingen → cwningod	rabbit → rabbits

Singular/Plural System:

bochdew → bochdewion	hamster → hamsters
twrch daear → tyrchod daea	mole → moles
chwistlen → chwistlod	shrew → shrews
gwiwer → gwiwerod	squirrel → squirrels
afanc → afancod	beaver → beavers
gwenci → gwencïod	weasel → weasels
ysgyfarnog → ysgyfarnogod	hare → hares
draenog → draenogod	hedgehog → hedgehogs
cath → cathod	cat→ cats
ci→ cŵn	dog→ dogs
dafad → defaid	sheep → sheeps
gafr → geifr	goat → goats
buwch → buchod	cow → cows
ceffyl→ ceffylau	horse → horses
llwynog → llwynogod (N)	fox → foxes
morlo → morloi	seal → seals
arth → eirth	bear → bears
blaidd → bleiddiaid	wolf → wolves
llew → llewod	lion → lions
teigr → teigrod	tiger → tigers
eliffant → eliffantod	elephant → elephants

Plantlife

Trees

Collective/Singulative System:

coed → coeden	trees → a tree
gwern→ gwernen	alders → an alder
onn → onnen	ashes → an ash
bedw→ bedwen	birches → a birch
ysgaw → ysgawen	elders → an elder
cyll→ collen	hazels → a hazel
celyn → celynnen	holly tree → a holly tree
derw → derwen	oaks → an oak
cerddin→ cerddinen	rowans → a rowan
helyg→ helygen	willows → a willow
sycamorwydd→ sycamorwydden	sycamores → a sycamore
castanwydd → castanwydden	chestnut trees → a chestnut tree
pinwydd → pinwydden	pine trees → a pine tree
ffynidwydd → ffynidwydden	fir trees → a fir tree
pefrwydd → pefrwydden	spruce trees → a spruce tree
cegiden → gegid	hemlock trees
cerddin → cerddinen	mountian ash
ffyllwydd → ffyllwydden	cypresses → cypress
gwern → gwernen	lime trees → lime tree
gwydd-wydd → gwydd-wydden	honeysuckles → honeysuckle
meryw → merywen	junipers → juniper
poplys → poplysen	poplars → poplar
yw → ywen	yews → yew

Flowers

Collective/Singulative System:

cennin pedr → cenhinen bedr	daffodils → a daffodil

bwtsias y gog → bwtsiasen y gog | bluebells → bluebell
gellesg → gellesgen | irises → an iris
rhos → rhosyn (Lit) ♦ | roses → a rose

Faux singulatives:
blodyn → blodau | flower → flowers
rhosyn → rhosod ♦ | rose → roses
briallen → briallu | primrose → primroses

Singular/plural system:
tiwlip → tiwlipau ♦ | tulip → tulips
tegeirian → tegeirianau | orchid → orchids
fioled → fioledau♦ | violet → violets
lili→ lilis ♦ | lily → lillies

Misc Collective Terms

plant → plentyn | child → children
sêr → seren | stars → a star
plu → pluen | feathers → a feather

A.2 Arabic Word List

Abbreviations: NA = North African, Yem = Yemeni Arabic, Lev = Levantine Arabic, Lib = libyan Arabic, Egy = Egyptian Arabic

Granular Materials

Collective/Singulative System

ghubaar → ghubaarah (غبارة ← غبار) dust → a speck of dust ♥
rimaal → rimaalah (رمالة ← رمال) sand → a grain of sand[22]
ḥaṣ → ḥaṣah (حصة ← حصى) gravel/pebbles → grain of gravel / pebble ♥

Food Items

Grains, Seeds & Pulses

Collective/Singulative System

sha3iir → sha3iirah (شعيْرة ← شعيْر) barley → a grain of barley ♥
badhr → bidhrah (بِذْرة ← بَذْ) seeds → a seed
dur → durah (ذرة ← ذر) corn → a piece of corn ♥
ḥumms̩ → ḥimṣah (حمصة ← حمص) chickpeas → a chickpea ♥
3das → 3dasah →(عدسة ← عدس) lentils → a lentil ♥
fuul → fuulah (فولة ← فول) beans → a bean
bun → bunnah (بنة ← بنّ) coffee beans → a coffee bean

Collective / classifier

shufaan → habah shufaan (حبة شوفان ← شوفان) oats → an oat ♣
baazla' → haba baazla"→(حبة بازلاء ← بازلاء) peas → a pea ♣

[22]c.f. singular/plural version رمال ← رمل

Nuts

Collective/Singulative System

mikssar → mikssarah (مكسر ← مكسَرة)	nuts → a nut[23]
lawz → lawzah (لوز ← لوزة)	almonds → an almond
jawz → jawzah (جوز ← جوزة)	walnut → a walnut ♥
jawz kaashuu → jawzat kaashuu (جَوز كاشو ← جَوزة كاشو)	cashew → cashews ♥♠
bunduq → bunduqah (بُنْدُق ← بُنْدُقَة)	hazelnut → hazelnuts
jawz albiqaan → jawzat albiqaan (جوز البيكان ← جوزة البيكان)	pecans → pecans ♥♠
fustuq → fustuqah (فُسْتُق ← فُسْتُقة)	pistachios/peanuts → pistachios/peanut[24]

Berries and aggregate fruit

Collective/Singulative System

tuut → tuutah →(تُوْت ← تُوْتَة)	berries → berry ♥
tuut brii→ tuut briyah (توت بري ← توتة برية)	a cranberry ♥♠
kashmish → kashmishah (كشمش ← كشمشة)	currents → current
zabiib → zabiibah →(زَبِيْب ← زَبِيْبَة)	raisins → raisin
3liiq → 3liiqah (عُلَّيْقة ← عليق)	raspberries/blackberries → raspberry/blackberry ♥
3nb → 3nbah (عِنَب ← غِنبةُ)	grapes → grape
faraawil → faraawilah (فراول ← فراولة)	strawberries → strawberry

Root Vegetables

Collective/Singulative System

fujl → fujlah (فُجْل ← فُجْلَة)	radishes → radish
shamandar → shamandarah (شَمَندَر ← شَمَندَرة)	beetroots → beetroot ♥

[23]Some speakers do not accept مِكَسَّر as a valid collective, preferring to use the sound plural and feminine singular مكسرة ← مكسرات

[24]The same term may be used for both pistachios and peanuts. In the sense of peanuts, the compound collective فول سوداني ← فولة سوداني 'fuul suudaanii → fuulat suudaanii' was proposed as an alternative, using the collective/singulative فول (*bean)* as the head noun.

jazar → jazrah (جَزَرَة ← جَزَر)	carrots → carrot
jazar a'byaḍ → jazrah beyḍa' (جزرة بيضاء ← جزر أبيض)	parsnips → a parsnip ♠[25]
laft → laftah (لفتة ← لفت)	turnips → turnip
baṣal 3sqlaanii → baṣalah 3sqlaanii (بَصَلَة عسقلاني←بصل عسقلاني)	shallots → shallot ♠
baṣal → baṣalah (بَصَلَة ← بَصَل)	onions → onion
kurraat → kurraatah (كُرَّاتة ← كُرَّات)	leeks → leek ♥

Cruciferous Vegetables

Collective/Singulative System

kurunb bruksil→ kurunbah bruksil (كُرْنُبة بروكسِل ← كُرُنُب بْروكْسِل)	sprouts → sprout[26]
khas → khasah (خَسَّة ← خَسّ)	lettuces → lettuce
kurunb → kurunbah (كُرْنُبَة ← كُرُنْب)	cabbages → a cabbage[27]
qurnabiiṭ → qurnabiiṭah (قرنبيطة ← قرنبيط)	cauliflower → cauliflowers
brukulii → brukuliyah (بْروكُلية ← بْروكُلي)	broccolis → broccoli ♥

Stone Fruits and vegetables

Collective/Singulative System

zaytuun → zaytuunah (زَيْتُوْنَة ← زَيْتُوْن)	olives → an olive
karaz → karazah (كرزة ← كرز)	cherries→ a cherry
tamr → tamrah (تمرة ← تمر)	dates → a date
moshosh → moshmoshah (مشمشة ← مشمش)	apricots → an apricot
barquuq → barquuqah (بَرْقُوْقَة ← بَرْقُوْق)	plums → a plum
khukh → khukhah (خوخة ← خوخ)	peaches → a peach
afukaadu → afukaaduah (أَفُوْكادُوْة ← أَفُوْكادُوْ)	avacados → an avocado ♦♣

Singular/plural system (contested)

niktaariin → niktaariinaat (نِكْتارينات→نِكْتارين)	nectarine → nectarines ♦ [28]

[25]Literally 'white carrot'. One speaker attested the alternative جزرة بري ← جزر بري, which still patterns as collective/singulative.

[26]Lit. 'brussel (borrowed) sprout (col)' One speaker attested براعم ← برعم which has a singular/plural contrast.

[27]One speaker suggested 'milfuuf' ملفوف as a the collective term.

[28]One speaker attested 'niktaariin' نِكْتارين as a collective form.

Other Fruits and Vegetables

Collective/Singulative System

fuṭr → fuṭrah (فُطْرَة ← فُطْر)	mushrooms → mushroom[29]
tiin → tiinah (تِيْنَة ← تِيْن)	figs → fig ♥
'ijaaṣ → ijaaṣah (إجاصة ← إجاص)	pears → pear[30]
kiiwii → kiiwiiah (كِيْوِية ← كِيْوِي)	kiwi → kiwi ♣♦
burtaqaal → burtaqaalah (برتقالة ← برتقال)	oranges → orange
leymuun → leymuunah (ليمونة ← ليمون)	lemons → lemon
tufaaḥ → tufaaḥah (تفاحة ← تفاح)	apples → apple
ṭamaaṭim → ṭamaaṭimah (طَمَاْطِمَة ← طَمَاْطِم)	tomato → tomatoes ♦♣
fulful → fulfulah (فُلْفُلَة ← فُلْفُل)	peppers → a pepper ♥
mawz → mawzah (موزة ← موز)	banana → bananas
rumaan → rumaanah (رمانة ← رمان)	pomegranates → a pomegranate
ananas → ananasah (أَنَانَاسة ← أَناناس)	pineapple → pineapples ♦♣
khiyaar → khyarah (خِيَاْرَة ← خِيَاْر)	cucumbers → cucumber
kusaa → kusa'ah (كُوْسَاْة ← كُوْسَا)	courgettes → courgette ♥
shamaam → shamaamah (شمامة ← شمام)	melons → melon
baṭikh → baṭikhah (بطيخة ← بطيخ)	watermelons → watermelon
qar3 → qar3ah (قَرْعَة) ← (قَرْع)	pumpkin → pumpkins ♥

Living Beings

Wingless Insects (including Arachnids and Molluscs):

Collective/Singulative System

qaml → qamlah (قَمْلَة ← قَمْلَ)	louse → a lice
naml → namlah (نَمْلَة ← نَمْل)	ants/mites → an ant/mite
duud → dudah (دُوْدَة ← دُوْد)	maggots/worms → a maggot/worm

29 One speaker suggested the singular/plural contrast فطور ← فطر

30 One speaker attested كمثرى as a collective term.

yaraqaan → yaraqanah (يَرَقَانَة ← يَرَقَان)	caterpillars → caterpillar[31]
qrad → qradah (قْرَادة ← قْرَاد)	ticks → a tick
ḥalazuun → ḥalazuunah (حَلَزُوْنَة ← حَلَزُوْن)	snails → a snail[32]
salaaj → salaajah (سلاجة ← سلاج)	slugs → a slug
buzziiq → buzziiqah (بُزَّيْقَة ← بُزَّيْق) (LA)	slugs → a slug

Singular/plural System

burghuth → baraghiith (بَرَاغِيْث ← بُرْغُوْث)	flea → fleas
3nkabuut → 3nakib (عَنَاكِب ← عَنْكَبُوْت)	spider → spiders
3qrab → 3qarib (عَقَارِب ← عَقْرَب)	scorpion → scorpions

Winged Insects:

Collective/Singulative System

dhubaab → dhubaabah (ذُبَابَة ← ذُبَاب)	flies → a fly
naḥl → naḥlah (نَحْلَة ← نَحْل)	bees → a bee
nu3uṣ → nu3uṣah (بُعُوْضَة ← بُعُوْض)	mosquitoes → a mosquito
faraash → faraashah (فَرَاشَة ← فَرَاش)	butterfly → butterflies[33]
zuqruṭ → zuqruṭah (زُقُرطَة ← زُقرُط) (Lev)	wasps → a wasp
zuzuzii → zuzuziiah (زْزُوْزِيَة ← زْزُوْزِي) (NA)	wasps → a wasp
jarad → jaradah (جْرَادَة ← جْرَاد) (NA, Yem)	grasshoppers → a grasshopper
jaraad → jaraadah (جَرَادَة ← جَرَاد)	locusts → a locust

Singular/plural System

dabbar → dabaabiir (دَبَابِيْر ← دَبُّوْر)	wasp → wasps
ṣirṣar → ṣaraṣiir (صَرَاصِير ← صرصار)	cockroach
khunfusa’ → khanaafis (خَنَافِس ← خُنْفُسَاْء)	beetle → beetles
jundub → janaadib (جَنَادِب ← جُنْدُب)	grasshoppers → a grasshopper

[31] One speaker suggested the singular/plural contrast يرقات الفراشة ← يرقة الفراشة

[32] One speaker wanted collective for both.

[33] One speaker attested singular/plural contrast فراشات ← فراشة. Note this singular is identical to the singulative form and the plural is identical to the sound plural of the singulative form, perhaps indicating that for this speaker that BUTTERFLY has left the collective/singulative class.

Crustaceans:

Collective/Singulative System

maḥaar → maḥaarah (مَحَارْ ← مَحَارَة)	oysters → an oyster♥
quraudis → quraydisah (قُرَيْدِس ← قُرَيْدِسَة)	shrimps/prawns → shrimp/prawn ♥
jimbirii → jimbiriiah (جَمْبَرِي ← جَمْبَرِيَاية) EGY	shrimps/prawns → a shrimp/prawn
saraṭaan → saraṭaanah (سَرَطَانْ ← سَرَطَانَة)	crab → crabs ♥
ḥiiliiaan → ḥiiliiaanah (حِيْلْيَان ← حِيْلْيَانَة) GULF	lobsters → a lobster
karkand → karkandah (كَرْكَنْد ← كَرْكَنْدة)	lobsters → lobsters ♥

Fish:

Collective/Singulative System

samak → samakah (سمك ← سَمَكَة)	fishes → a fish
salimuun → salimuunah (سلمون ← سلمونة)	salmon → salmons ♥ [34]
samak alfaraz → samakah alfaraz (سمك الفرخ ← سمكة الفرخ)	perch → perches
a'nshuufah → a'nshuufah (أنشوف ← أنشوفة)	anchovy → anchovies [35])
samak albuqla → samakah albuqlah (البَقَلَة سمك ← البَقَلَة سمكة)	cod → cods
samak alshuuṭ → samakah alshuuṭ (سمكة الشبوط ← سمك الشبوط)	carp → carps

Singular/plural System

th3baan albḥr → th3lbiin albḥr (ثعبان البحر ← ثعابين البحر)	eel → eels

Amphibians and Reptiles

Collective/Singulative System:

br3ṭ → br3ṭah (بُرْعُط ← بُرْعُطة)	tadpoles → tadpole[36])
saḥaalii ma'iyah → saḥaaliyah ma'iyah (سحالي مائية ← سحلية سحلية)	newt → newts
tiras → tirasah (تِرَس ← تِرْسَة)	turtles → a turtle ♥
salaḥif → salaḥifah (سَلَاْحِف ← سُلَحْفَاْة)	tortoises → a tortoise

[34] One speaker suggested a compound version (سَمَك) سُلَيْمان ← (سَمَكَة) سُلَيْمان

[35] One speaker suggested a compound version الأنشوفة سمكة ← الأنشوفة سمك

[36] One person suggested the singular/plural contrast شرغوف ← شراغيف

Collective Only

iighwaanaa (إيغوانا) iquana → iguanas ♥[37]

Singular/Plural System:

zaaḥif → zuwaaḥif (زَاْحِف ← زَوَاْحِف) reptile → reptiles
th3baan → th3abiin (ثُعْبَاْن ← ثَعابين) snake → snakes
ṣifdi3 → ṣifadi3 (ضِفْدِع ← ضَفَاْدِع) frog/toad → frogs/toads
samandal → samaandal (سمندل ← سماندل) salamandar → salamanders

Birds:

Collective/Singulative System

ṭa'iir → ṭa'iirah (طَيْر ← طَيْرَة) birds → bird[38]
ḥamaam → ḥamaamah (حمام ← حمامة) pigeon → an pigeon
buum → buumah (بوم ← بومة) owls → an owl
baj3 → baj3ah (بجع ← بجعة) swans → a swan
ḥamal → ḥamalah (حَجَل ← حَجَلَة) partidge → partidges ♥
baṭ → baṭah (بط ← بطة) ducks → duck
dajaaj → dajaajah (دجاج ← دجاجة) chicken → chickens
iwaz → iwazah (إِوَزّ ← إِوَزّة) goose → geese

Singular/plural System

3ṣfur → 3ṣaafiir (عصفور ← عصافير) sparrows → a sparrow
3q3q → 3qaaq (عقعق ← عقاق) magpie → magpies
ghuraab → ghirbaan (غراب ← غربان) raven/crow → ravens/crows
nasr → nusuur (نَسْر ← نُسُوْر) vulture/eagle → vultures/eagles
diik → duyuuk (دِيْك ← دُيُوك) rooster → roosters

---

[37]One speaker attested this is a collective only, with no singulative form.
[38]One speaker prefered the singular/plural pair طائر ← طيور

Mammals:

Collective/Singulative System

fa'r → fa'rah (فَأْر ← فَأْرَة)	mice → a mouse[39]
zabab → zababah زباب ← زبابة	shrew → shrews
n3aaj → n3jah (نعاج ← نعجة)	ewes → an ewe
baqar → baqarah (بَقَر ← بقرة)	cow → cows

Singular/plural System

ghariir → ghararaan (غرير ← غرران)	badgers → badger
juraḍ → jirḍaan (جرذ ← جرذان)	rats → rat
khalad ala'rḍ → khaladaan ala'rḍ (خلد الأرض ← خلدان الأرض)	mole → moles
sinjaab → sinaajib (سنجاب ← سناجب)	squirrel → squirrels
arnib → araanib (أرنب ← أرانب)	rabbit/hare → rabbits/hares
qunfud → qanaafid (قنفذ ← قنافذ)	hedgehog → hedgehogs
qandis → qanaadis (قندس ← قنادس)	beaver → beavers
khinziir → khanaaziir (خنزير ← خنازير)	pig → pigs
kharuuf → khirfaan (خروف ← خراف)	sheep → sheeps
maa3iz → mawaa3iz (ماعز ← مواعز)	goat → goats
ḥiṣaan → ḥuṣun (حصان→ حصن)	horse → horses
khuffaash → khafaafiish (خفاش ← خفافيش)	bat → bats
qiṭ → qiṭaṭ (قطط ← قطّة)	cat→ cats
kalb → kilaab (كلاب ← كلب)	dog→ dogs
tha3lab → tha3laab (ثعلب→ ثعالب)	fox → foxes
dubb → dibabah (دبّ ← دببة)	bear → bears
dhi'b → dhi'aab (ذئب ← ذئاب)	wolf → wolves
asad → usuud (أَسَد ← أُسُود)	lion → lions
namir → numuur (نمر ← نمور)	tiger → tigers
fiil → fiyalah (فيل ← فيلة)	elephant → elephants

[39] One speaker attested the preference of singular/plural pair فأر ← فئران

Plantlife

Trees

For the tree terms, there is a preference for some speakers to compound some tree terms with 'shajar' (tree) as the head noun in a collective/singulative contrast. I've indicated where this is the case with ✿

Collective/Singulative System

shajar → shajarah (شَجَرة ← شَجَ)	trees → a tree
ṣafṣaaf → ṣafṣaafah (صَفْصَافَة ← صَفْصَاف)	willows → a willow ✿
bunduq → bunduqah (بُنْدُقَة ←بُنْدُق)	hazel trees → a hazel tree[40]
zaytuun → zaytuunah (زَيْتُوْنَة ←زَيْتُوْن)	Olive trees → Olive tree ✿[41]
sadr → sadrah (سِدْرَة ← سِدْر)	lotus trees → lotus tree[42]
ṭalḥ → ṭalḥah (طَلْحَة ← طَلْح)	banana trees → a banana tree[43]
ghaar → ghaarah (غَارَة ← غَار)	laurel trees → laurel tree ✿
mishmish → mishmishah (مِشْمِشَة ← مِشْمِش)	apricot trees → apricot tree ✿[44]
nakhl → nakhlah (نَخلِة ← نَخِل)	palm trees → palm tree ✿ [45]
arz → arzah (أرزة ← أرز)	cedar → cedars
sarw → sarwah (سَرْوة ← سَرْو)	cyprusses → cypruss
kharrub → kharrubah (خَرُّوبٌة ← خَرُّوبٌ)	carob trees → carob tree
shajar alban → shajarah alban (شجرة البن ← شجر البن)	ben trees → ben tree

[40]Lit. 'hazelnut'.
[41]Lit. 'olive'.
[42]One speaker suggested an alternative, with shajar as the head noun شجرة اللوتس ← شجر اللوتس
[43]Some speakers prefer an alternative, with shajar as the head noun شجرة الموز ← شجر الموز
[44]Lit. 'apricot"
[45]Lit. 'palms'

Flowers

Collective/Singulative System

zahr → zahrah (زهرة ← زهر) flowers → a flower
zanbaq → zanbaqah (زَنْبَقَة ← زَنْبَق) lillies → a lilly[46]
ward → wardah (وَرْدَة ← وَرْد) rose → roses
banafsaj → banafsajah (بَنَفْسَجَة ← بَنَفْسَج) violets → a violet[47]
khashkhaash → khashkhaashah (خَشْخَاْشَة ← خَشْخَاْش) poppies → a poppy ♥
zahr altuuliib → zahrah altuuliib (زهرة التُوْلِيْب ← زهر التُوْلِيْب) tulips → tulip
zahr aluurkiid → zahrah aluurkiidah (زهرة الأورْكيد← زهر الاوركيدة) tulips → tulip

Misc

bayḍ → bayḍah (بيضة ← بيض) eggs → an egg

[46]One speaker attested a singular/plural contrast زنابق ← زنبق
[47]For some speakers, the contrast is singular/plural بنافسج ← بنفسج