Unsupervised Machine Translation

Mikel Artetxe

IXA NLP group – University of the Basque Country (UPV/EHU)

Unsupervised machine translation

Unsupervised machine translation

Conventional (supervised) machine translation

Unsupervised machine translation

Conventional (supervised) machine translation

stneilc od ton lles slacituecamrahp ni eporue

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	

Parallel corpus (translation examples)

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . ??? sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc $\boxed{\text{od}}$ ton lles slacituecamrahp ni eporue
saphsin	neilcset

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset on

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset on nevned

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset on nevned

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset on nevned emidicsan ne ueorap

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset on nevned emidicsan ne ueorap

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset on nevned emidicsan ne ueorap

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset on nevned emidicsan ne ueorap

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset on nevned emidicsan ne ueorap

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

hsilgen	stneilc od ton lles slacituecamrahp ni eporue
saphsin	neilcset on nevned emidicsan ne ueorap

Parallel corpus (translation examples)	
aicrag dna setaicossa . ragica y aosaicsod .	eht stneilc dna eht setaicossa era seimene . sol neilcset y sol aosaicsod nos eenimsog .
solrac aicrag sah eerht setaicossa . racsol ragica eiten sert aosaicsod .	eht ynapmoc sah eerht spuorg . al meerpas eiten sert urgsop .
sih setaicossa era ton gnorts . sus aosaicsod on nos reufset .	sti spuorg era ni eporue . sus urgsop senat ne ueorap .
aicrag sah a ynapmoc osla . ragica matneib eiten uan meerpas .	eht nredom spuorg lles gnorts slacituecamrahp . sol urgsop omredson nevned emidicsan reufset .
sti stneilc era yrgna . sus neilcset senat neafadsod .	eht spuorg od ton lles eninaznez . sol urgsop on nevned nazazinan .
eht setaicossa era osla yrgna . sol aosaicsod matneib senat neafadsod .	eht llams spuorg era ton nredom . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

english	clients do not sell pharmaceuticals in europe
saphsin	neilcset on nevned emidicsan ne ueorap

Parallel corpus (translation examples)	
garcia and associates . ragica y aosaicsod .	the clients and the associates are enemies . sol neilcset y sol aosaicsod nos eenimsog .
carlos garcia has three associates . racsol ragica eiten sert aosaicsod .	the company has three groups . al meerpas eiten sert urgsop .
his associates are not strong . sus aosaicsod on nos reufset .	its groups are in europe . sus urgsop senat ne ueorap .
garcia has a company also . ragica matneib eiten uan meerpas .	the modern groups sell strong pharmaceuticals . sol urgsop omredson nevned emidicsan reufset .
its clients are angry . sus neilcset senat neafadsod .	the groups do not sell zenzanine . sol urgsop on nevned nazazinan .
the associates are also angry . sol aosaicsod matneib senat neafadsod .	the small groups are not modern . sol urgsop epeuqsoñ on nos omredson .

*Inspired by Knight (1997)

Unsupervised machine translation

Conventional (supervised) machine translation

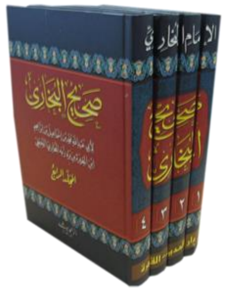
english	clients do not sell pharmaceuticals in europe
spanish	clientes no venden medicinas en europa

Parallel corpus (translation examples)	
garcia and associates . garcia y asociados .	the clients and the associates are enemies . los clientes y los asociados son enemigos .
carlos garcia has three associates . carlos garcia tiene tres asociados .	the company has three groups . la empresa tiene tres grupos .
his associates are not strong . sus asociados no son fuertes .	its groups are in europe . sus grupos estan en europa .
garcia has a company also . garcia tambien tiene una empresa .	the modern groups sell strong pharmaceuticals . los grupos modernos venden medicinas fuertes .
its clients are angry . sus clientes estan enfadados .	the groups do not sell zenzanine . los grupos no venden zanzanina .
the associates are also angry . los asociados tambien estan enfadados .	the small groups are not modern . los grupos pequeños no son modernos .

*Inspired by Knight (1997)

Unsupervised machine translation

Arabic corpus



Conventional (supervised) machine translation

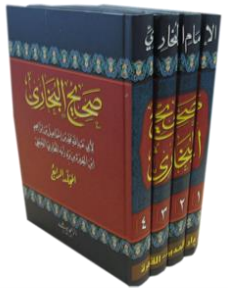
english	clients do not sell pharmaceuticals in europe
spanish	clientes no venden medicinas en europa

Parallel corpus (translation examples)	
garcia and associates . garcia y asociados .	the clients and the associates are enemies . los clientes y los asociados son enemigos .
carlos garcia has three associates . carlos garcia tiene tres asociados .	the company has three groups . la empresa tiene tres grupos .
his associates are not strong . sus asociados no son fuertes .	its groups are in europe . sus grupos estan en europa .
garcia has a company also . garcia tambien tiene una empresa .	the modern groups sell strong pharmaceuticals . los grupos modernos venden medicinas fuertes .
its clients are angry . sus clientes estan enfadados .	the groups do not sell zenzanine . los grupos no venden zanzanina .
the associates are also angry . los asociados tambien estan enfadados .	the small groups are not modern . los grupos pequeños no son modernos .

*Inspired by Knight (1997)

Unsupervised machine translation

Arabic corpus



Chinese corpus

Conventional (supervised) machine translation

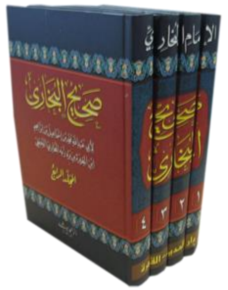
english	clients do not sell pharmaceuticals in europe
spanish	clientes no venden medicinas en europa

Parallel corpus (translation examples)	
garcia and associates . garcia y asociados .	the clients and the associates are enemies . los clientes y los asociados son enemigos .
carlos garcia has three associates . carlos garcia tiene tres asociados .	the company has three groups . la empresa tiene tres grupos .
his associates are not strong . sus asociados no son fuertes .	its groups are in europe . sus grupos estan en europa .
garcia has a company also . garcia tambien tiene una empresa .	the modern groups sell strong pharmaceuticals . los grupos modernos venden medicinas fuertes .
its clients are angry . sus clientes estan enfadados .	the groups do not sell zenzanine . los grupos no venden zanzanina .
the associates are also angry . los asociados tambien estan enfadados .	the small groups are not modern . los grupos pequeños no son modernos .

*Inspired by Knight (1997)

Unsupervised machine translation

Arabic corpus | **Chinese corpus**



non-parallel

Conventional (supervised) machine translation

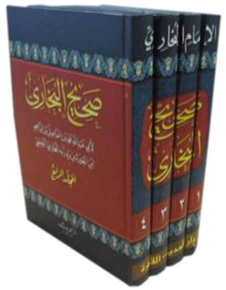
english	clients do not sell pharmaceuticals in europe
spanish	clientes no venden medicinas en europa

Parallel corpus (translation examples)	
garcia and associates . garcia y asociados .	the clients and the associates are enemies . los clientes y los asociados son enemigos .
carlos garcia has three associates . carlos garcia tiene tres asociados .	the company has three groups . la empresa tiene tres grupos .
his associates are not strong . sus asociados no son fuertes .	its groups are in europe . sus grupos estan en europa .
garcia has a company also . garcia tambien tiene una empresa .	the modern groups sell strong pharmaceuticals . los grupos modernos venden medicinas fuertes .
its clients are angry . sus clientes estan enfadados .	the groups do not sell zenzanine . los grupos no venden zanzanina .
the associates are also angry . los asociados tambien estan enfadados .	the small groups are not modern . los grupos pequeños no son modernos .

*Inspired by Knight (1997)

Unsupervised machine translation

Arabic corpus



non-parallel

Chinese corpus

Conventional (supervised) machine translation

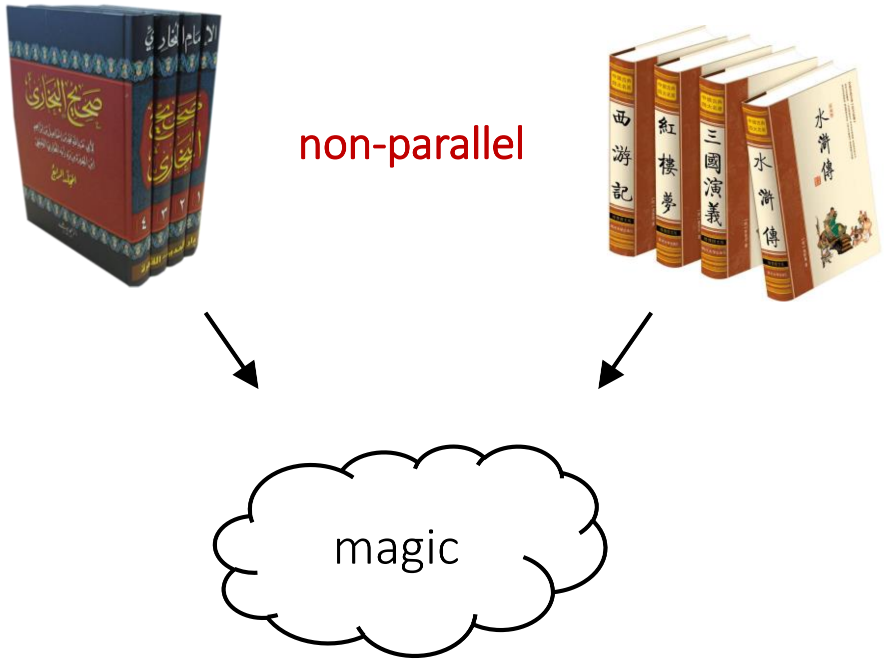
english	clients do not sell pharmaceuticals in europe
spanish	clientes no venden medicinas en europa

Parallel corpus (translation examples)	
garcia and associates . garcia y asociados .	the clients and the associates are enemies . los clientes y los asociados son enemigos .
carlos garcia has three associates . carlos garcia tiene tres asociados .	the company has three groups . la empresa tiene tres grupos .
his associates are not strong . sus asociados no son fuertes .	its groups are in europe . sus grupos estan en europa .
garcia has a company also . garcia tambien tiene una empresa .	the modern groups sell strong pharmaceuticals . los grupos modernos venden medicinas fuertes .
its clients are angry . sus clientes estan enfadados .	the groups do not sell zenzanine . los grupos no venden zanzanina .
the associates are also angry . los asociados tambien estan enfadados .	the small groups are not modern . los grupos pequeños no son modernos .

*Inspired by Knight (1997)

Unsupervised machine translation

Arabic corpus | **Chinese corpus**



Conventional (supervised) machine translation

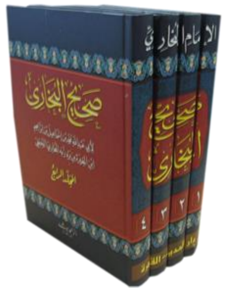
english	clients do not sell pharmaceuticals in europe
spanish	clientes no venden medicinas en europa

Parallel corpus (translation examples)	
garcia and associates . garcia y asociados .	the clients and the associates are enemies . los clientes y los asociados son enemigos .
carlos garcia has three associates . carlos garcia tiene tres asociados .	the company has three groups . la empresa tiene tres grupos .
his associates are not strong . sus asociados no son fuertes .	its groups are in europe . sus grupos estan en europa .
garcia has a company also . garcia tambien tiene una empresa .	the modern groups sell strong pharmaceuticals . los grupos modernos venden medicinas fuertes .
its clients are angry . sus clientes estan enfadados .	the groups do not sell zenzanine . los grupos no venden zanzanina .
the associates are also angry . los asociados tambien estan enfadados .	the small groups are not modern . los grupos pequeños no son modernos .

*Inspired by Knight (1997)

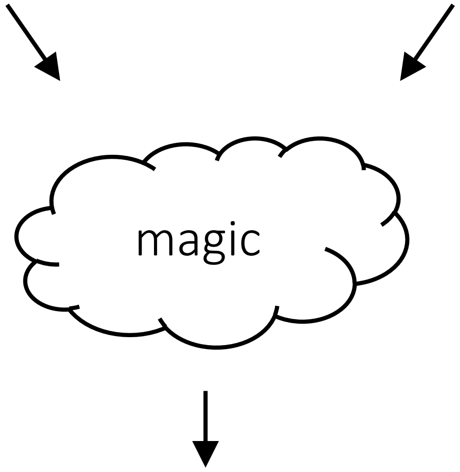
Unsupervised machine translation

Arabic corpus



non-parallel

Chinese corpus



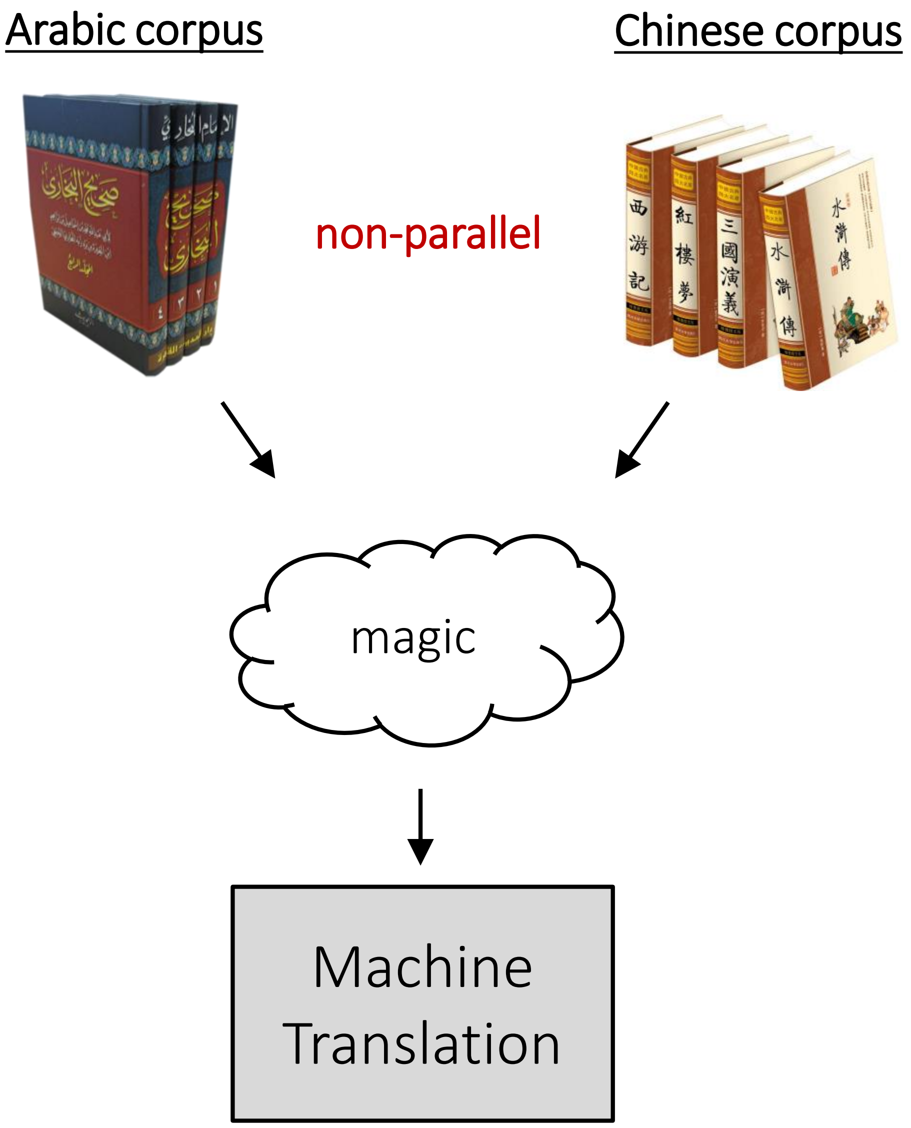
Conventional (supervised) machine translation

english	clients do not sell pharmaceuticals in europe
spanish	clientes no venden medicinas en europa

Parallel corpus (translation examples)	
garcia and associates . garcia y asociados .	the clients and the associates are enemies . los clientes y los asociados son enemigos .
carlos garcia has three associates . carlos garcia tiene tres asociados .	the company has three groups . la empresa tiene tres grupos .
his associates are not strong . sus asociados no son fuertes .	its groups are in europe . sus grupos estan en europa .
garcia has a company also . garcia tambien tiene una empresa .	the modern groups sell strong pharmaceuticals . los grupos modernos venden medicinas fuertes .
its clients are angry . sus clientes estan enfadados .	the groups do not sell zenzanine . los grupos no venden zanzanina .
the associates are also angry . los asociados tambien estan enfadados .	the small groups are not modern . los grupos pequeños no son modernos .

*Inspired by Knight (1997)

Unsupervised machine translation



Conventional (supervised) machine translation

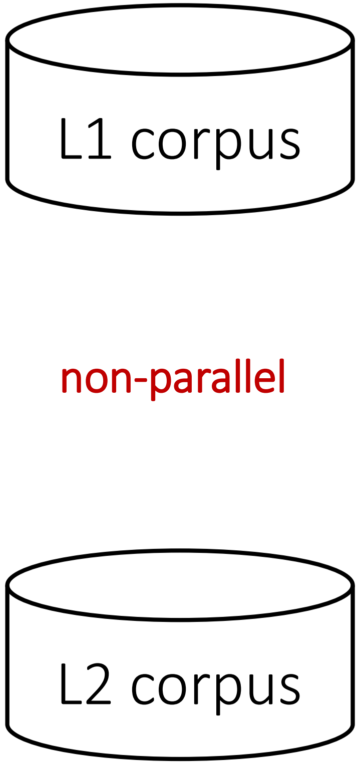
english	clients do not sell pharmaceuticals in europe
spanish	clientes no venden medicinas en europa

Parallel corpus (translation examples)	
garcia and associates . garcia y asociados .	the clients and the associates are enemies . los clientes y los asociados son enemigos .
carlos garcia has three associates . carlos garcia tiene tres asociados .	the company has three groups . la empresa tiene tres grupos .
his associates are not strong . sus asociados no son fuertes .	its groups are in europe . sus grupos estan en europa .
garcia has a company also . garcia tambien tiene una empresa .	the modern groups sell strong pharmaceuticals . los grupos modernos venden medicinas fuertes .
its clients are angry . sus clientes estan enfadados .	the groups do not sell zenzanine . los grupos no venden zanzanina .
the associates are also angry . los asociados tambien estan enfadados .	the small groups are not modern . los grupos pequeños no son modernos .

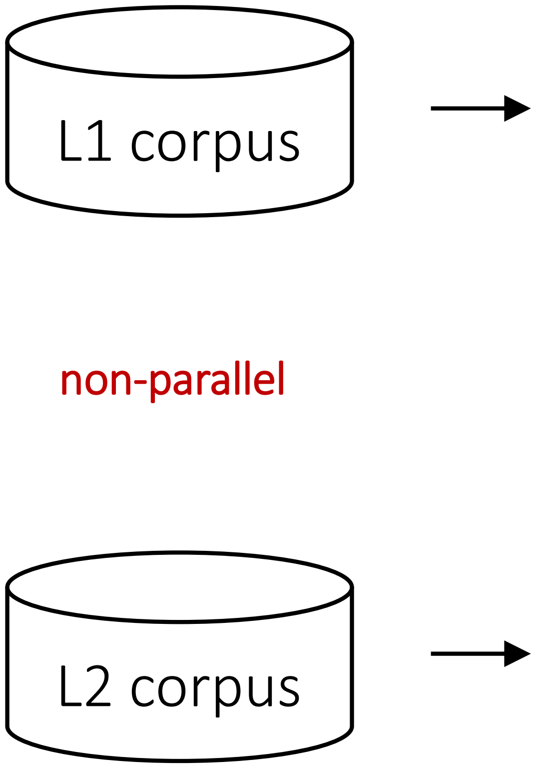
*Inspired by Knight (1997)

Outline

Outline

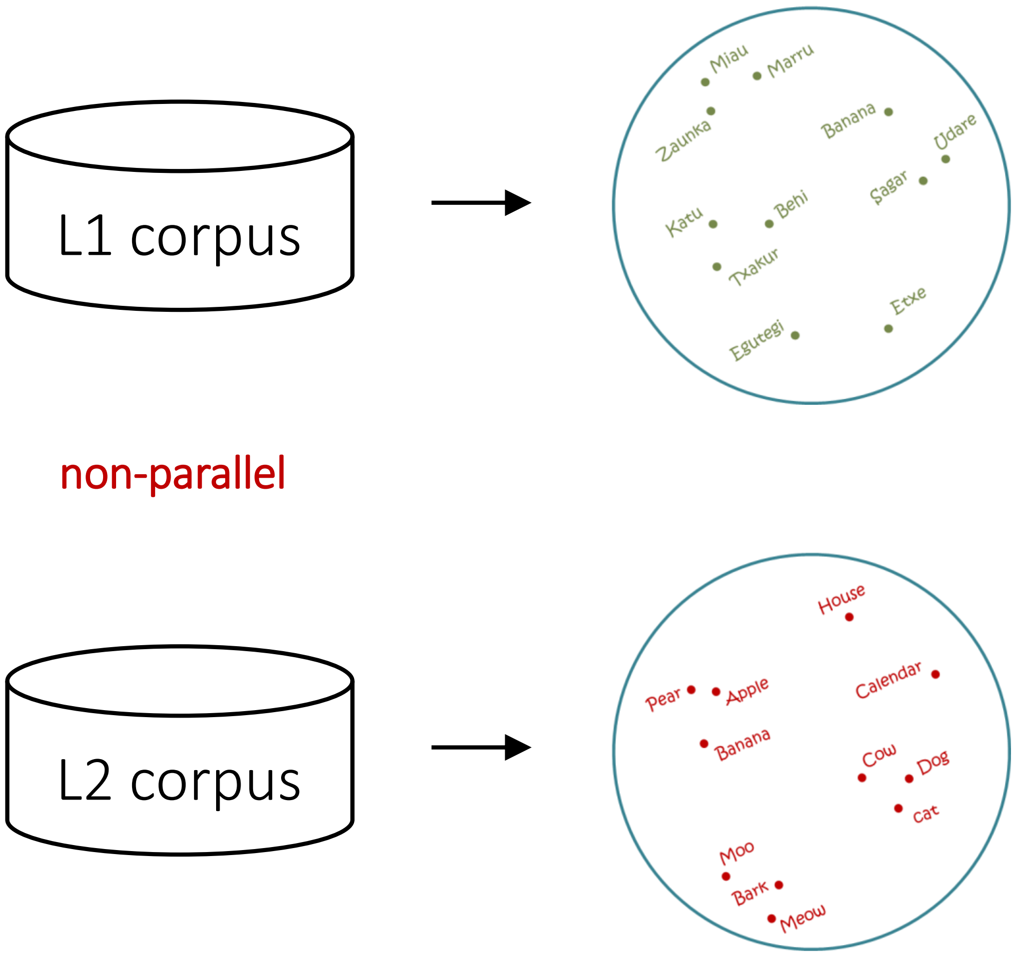


Outline



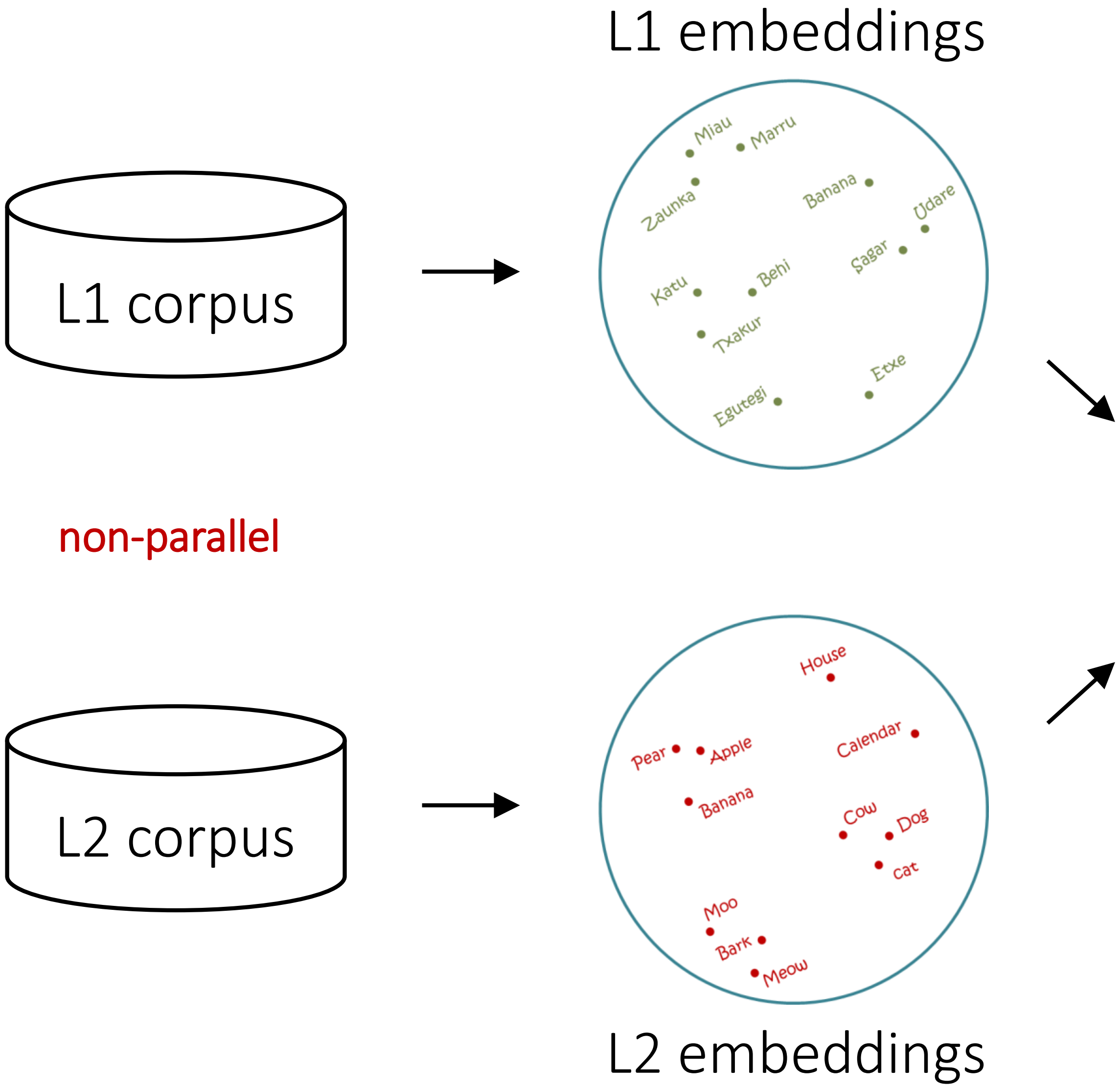
Outline

L1 embeddings

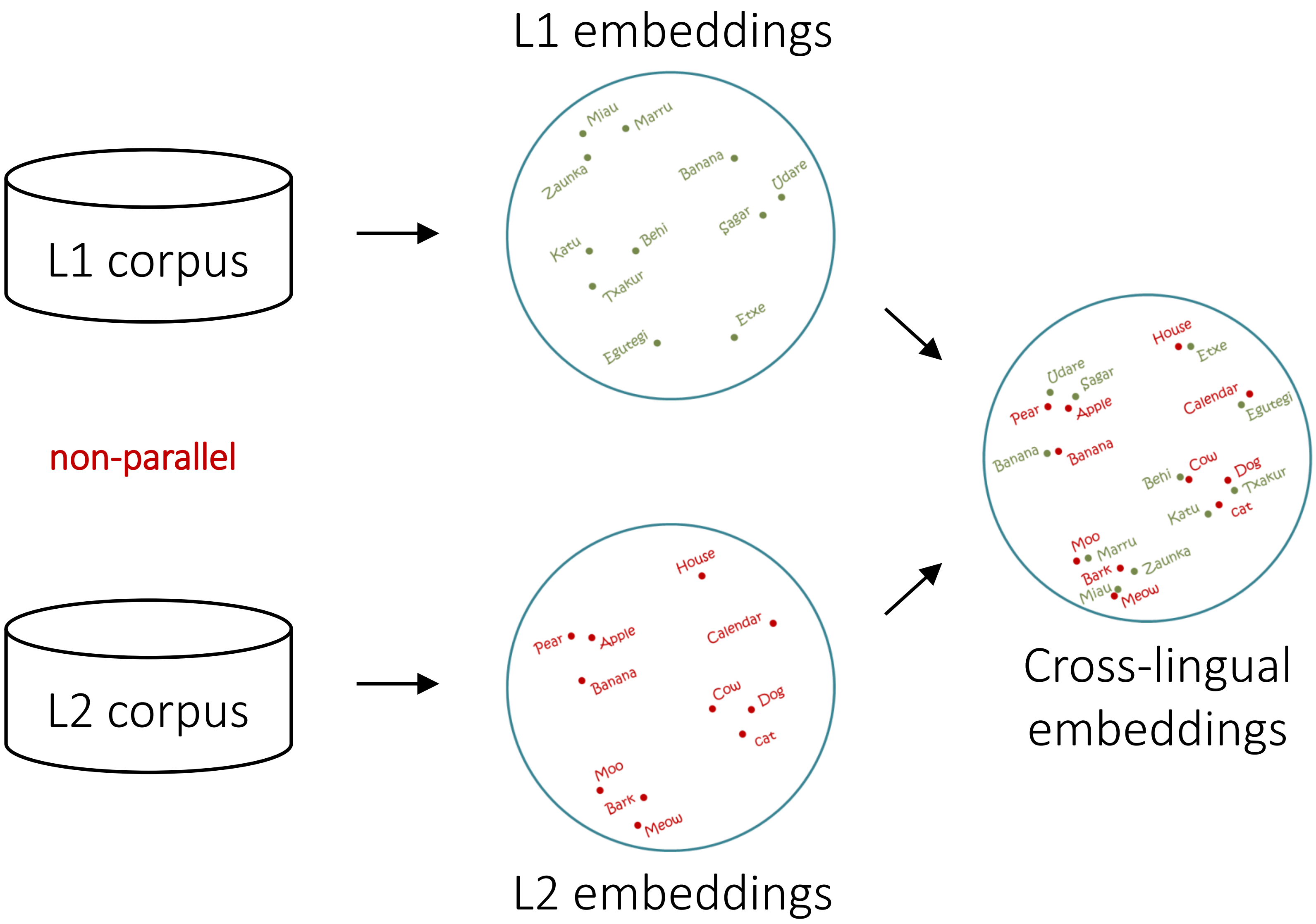


L2 embeddings

Outline

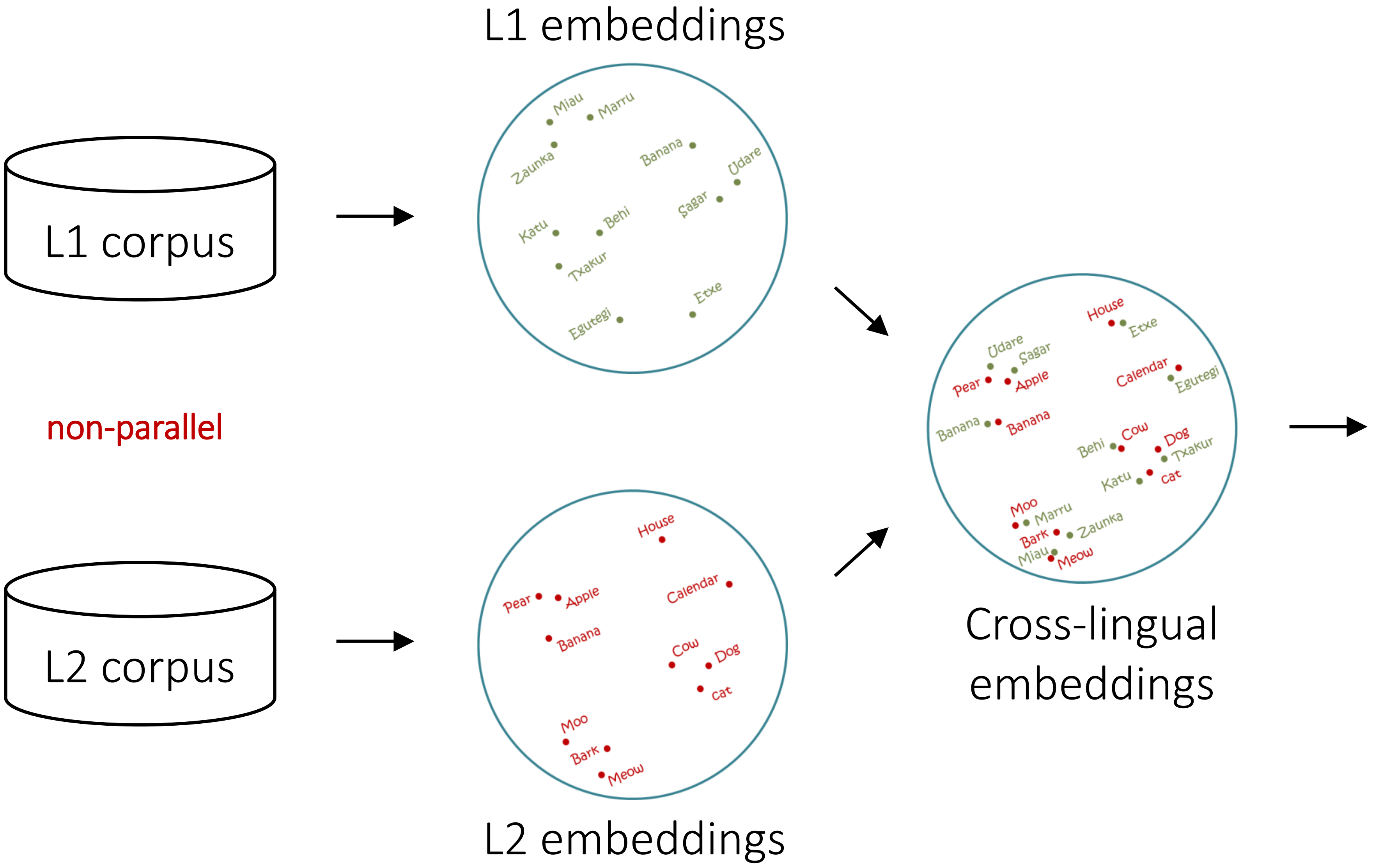

L1 embeddings
L1 corpus
Miau
Marru
Zaunka
Banana
Udare
Katu
Behi
Sagar
Txakur
Etxe
Egutegi
non-parallel
L2 corpus
House
Pear
Apple
Calendar
Banana
Cow
Dog
Cat
Moo
Bark
Meow
L2 embeddings

Outline

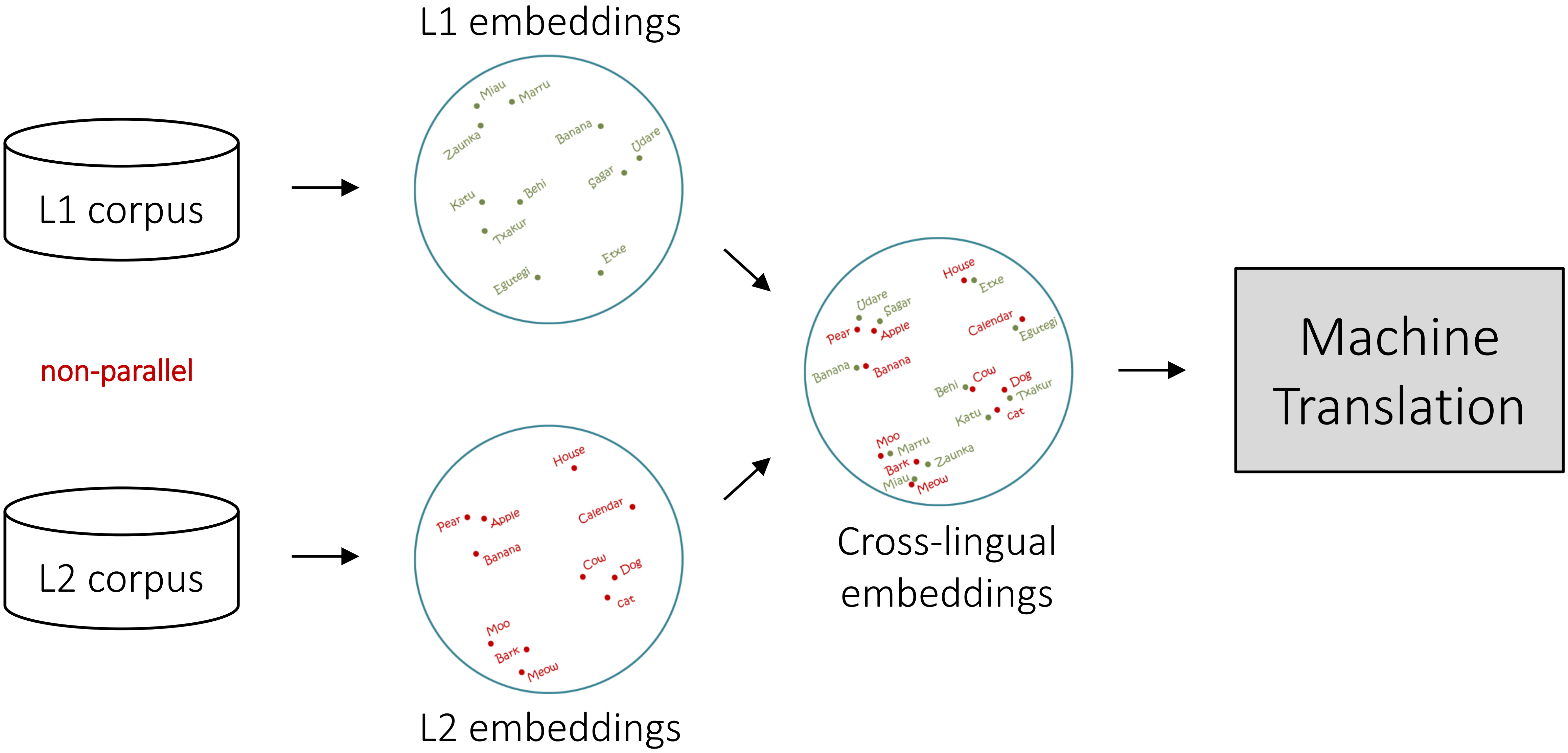


L1 embeddings
L1 corpus
non-parallel
L2 corpus
L2 embeddings
Cross-lingual embeddings
Miau
Marru
Zaunka
Banana
Udare
Katu
Behi
Sagar
Txakur
Etxe
Egutegi
House
Pear
Apple
Calendar
Cow
Dog
Cat
Moo
Bark
Meow

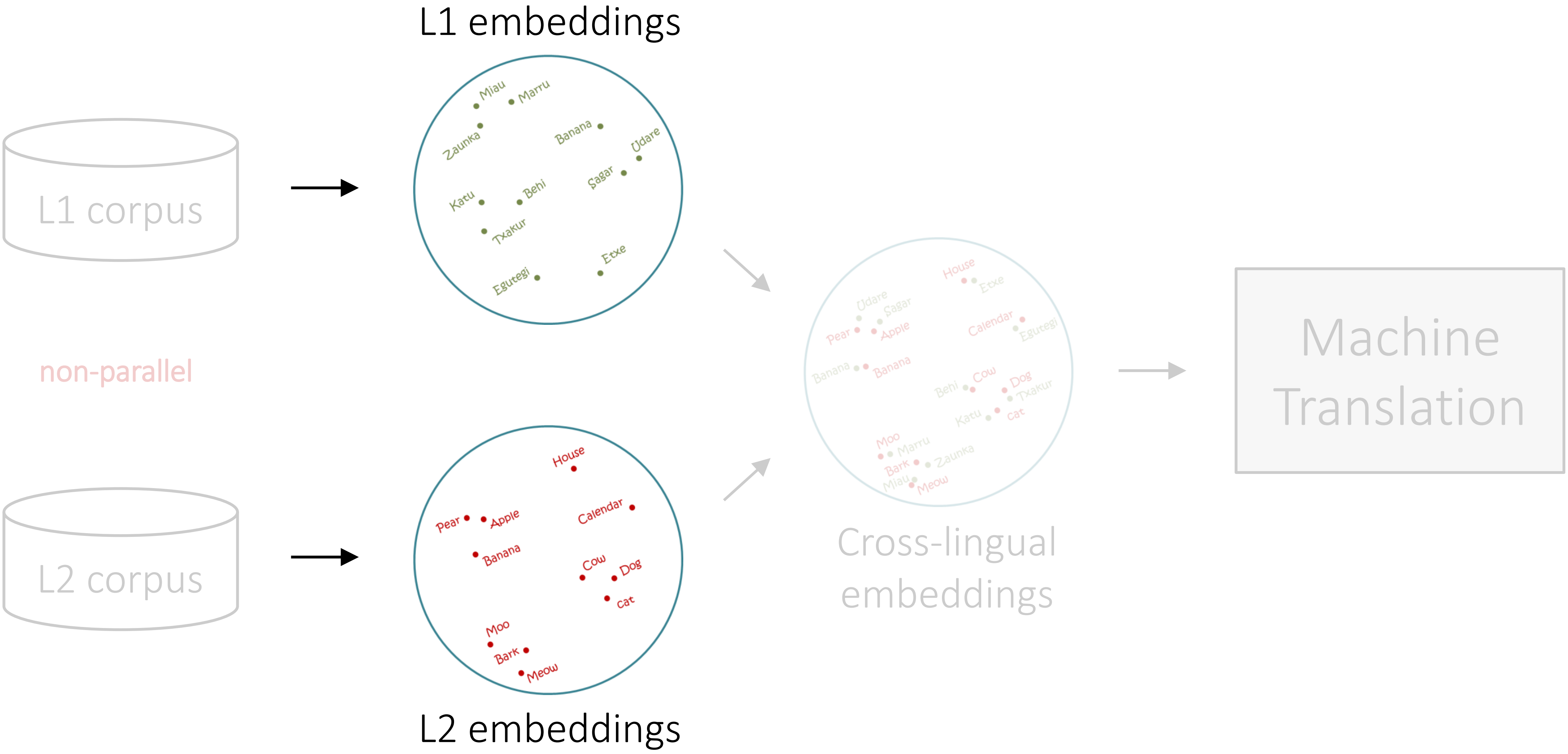
Outline


L1 embeddings
Miau
Marru
Zaunka
Banana
Udare
Katu
Behi
Sagar
Txakur
Etxe
Egutegi
L1 corpus
non-parallel
L2 corpus
House
Pear
Apple
Calendar
Banana
Cow
Dog
Cat
Moo
Bark
Meow
L2 embeddings
Cross-lingual
embeddings

Outline


L1 embeddings
L1 corpus
Miau
Marru
Zaunka
Banana
Udare
Katu
Behi
Sagar
Txakur
Etxe
Egutegi
non-parallel
L2 corpus
House
Pear
Apple
Calendar
Banana
Cow
Dog
Cat
Moo
Bark
Meow
L2 embeddings
Cross-lingual embeddings
Machine Translation

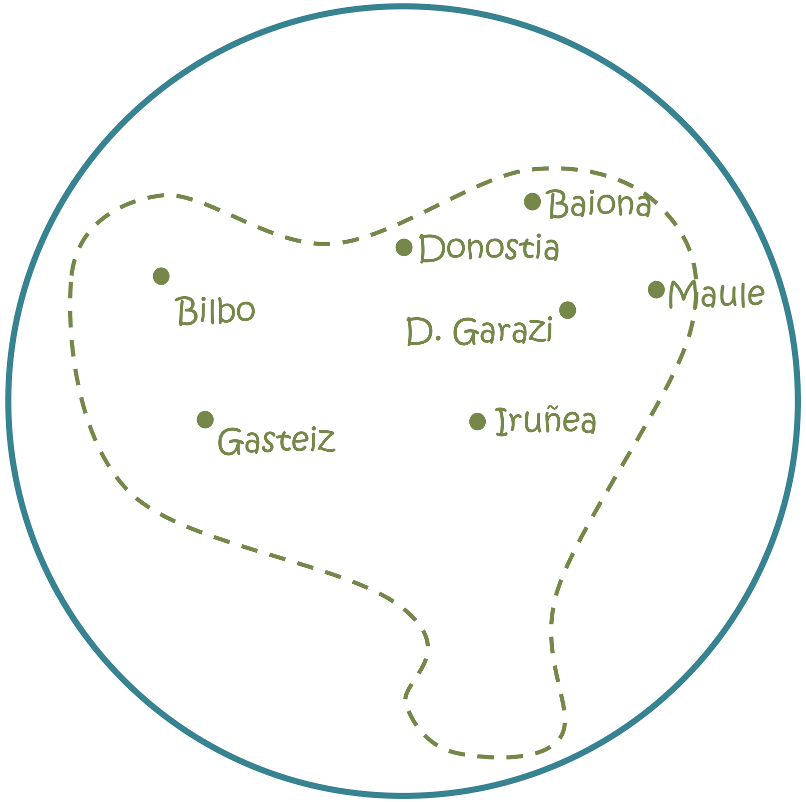
Outline



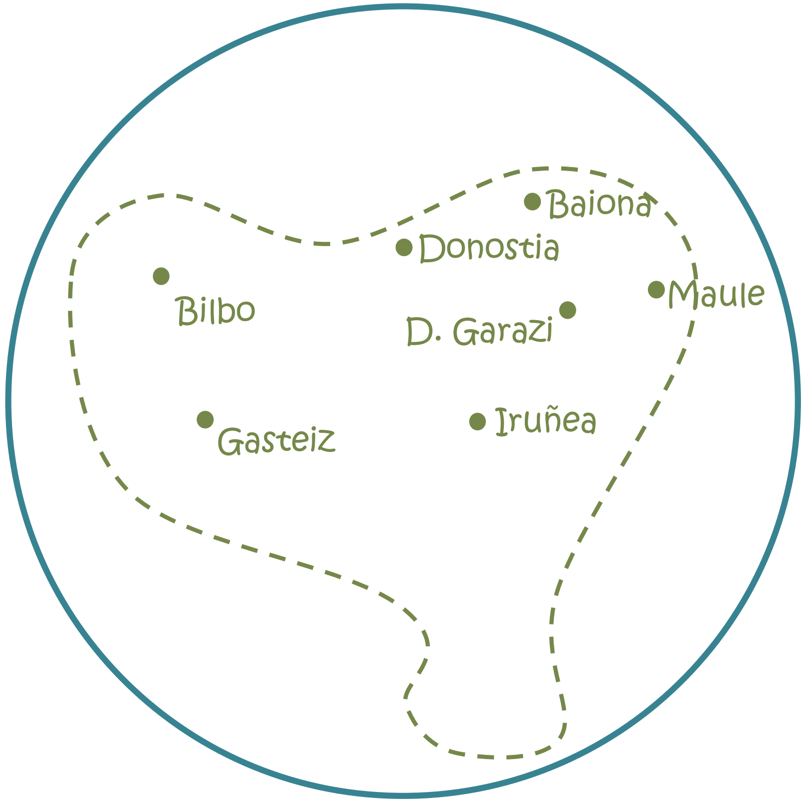
L1 embeddings
L1 corpus
Miau
Marru
Zaunka
Banana
Udare
Katu
Behi
Sagar
Txakur
Etxe
Egutegi
non-parallel
L2 corpus
House
Pear
Apple
Calendar
Banana
Cow
Dog
cat
Moo
Bark
Meow
L2 embeddings
Cross-lingual
embeddings
Machine
Translation

Embeddings

Embeddings

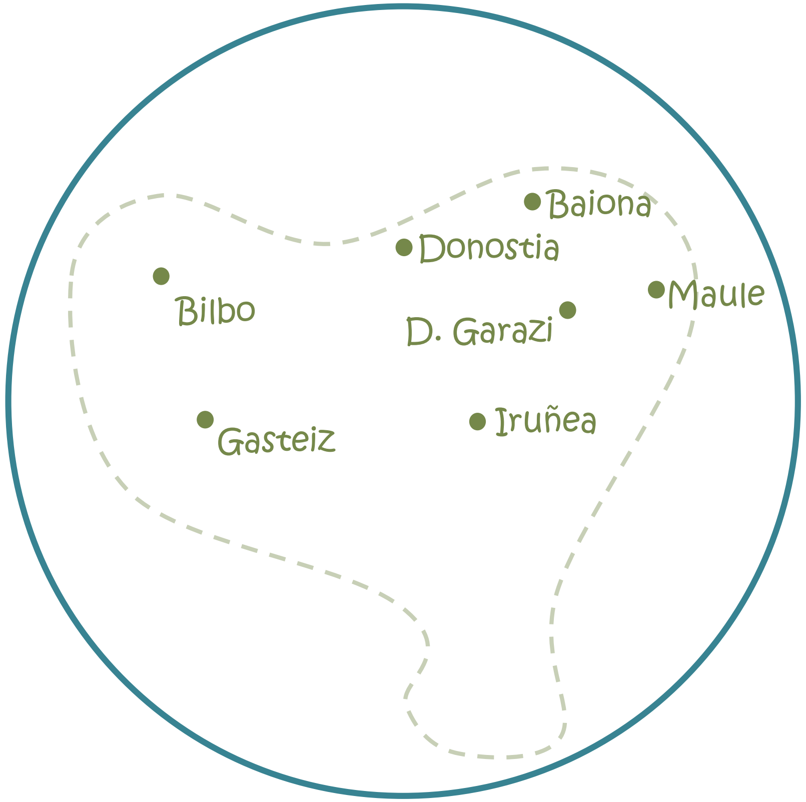


Embeddings



Geographical space

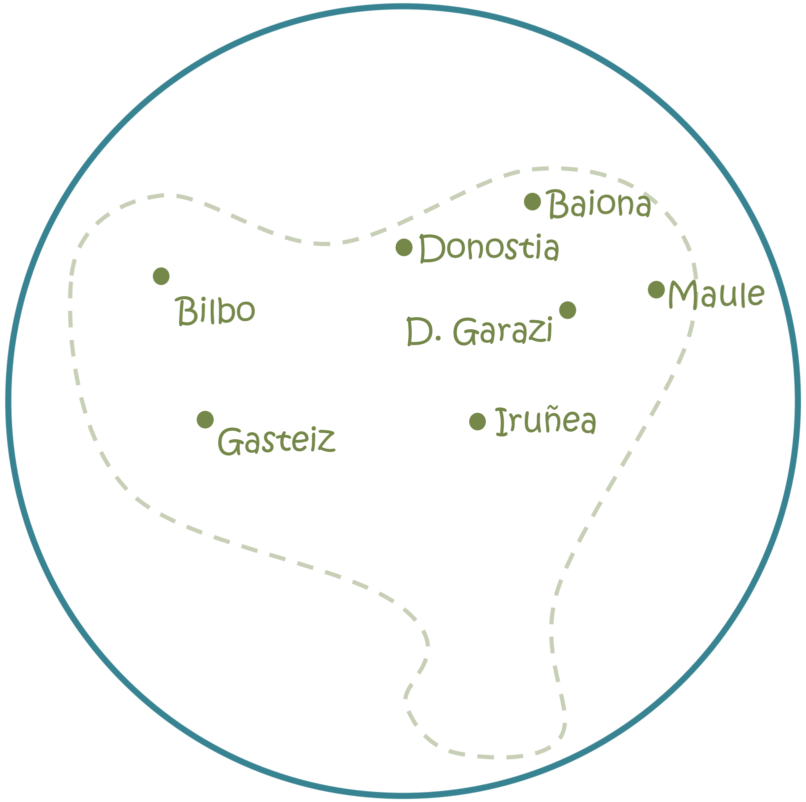
Embeddings



<u>Geographical space</u>

- Cities

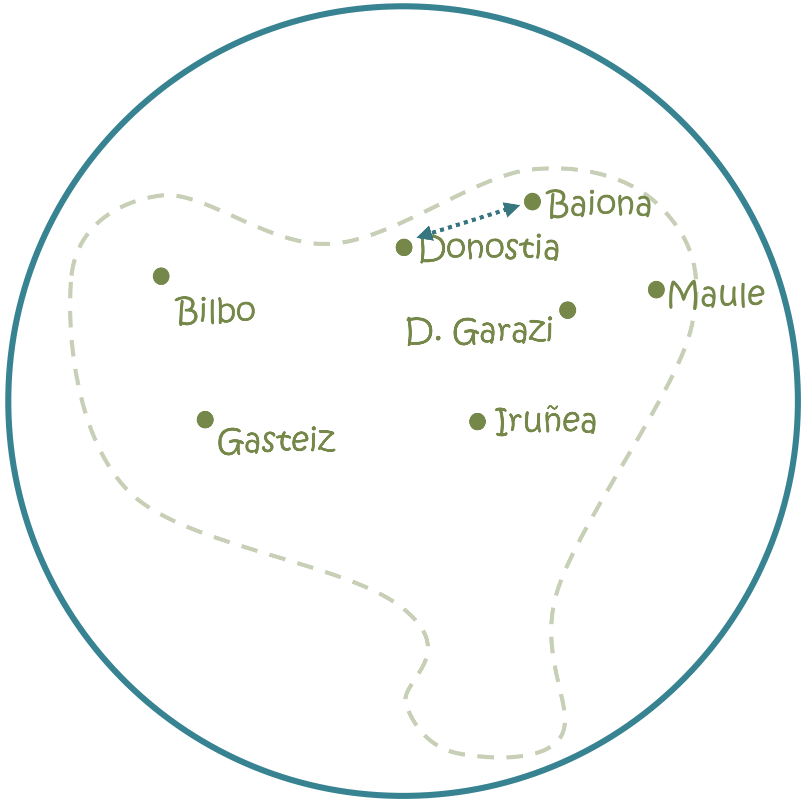
Embeddings



Geographical space

- Cities
- Meaningful distances

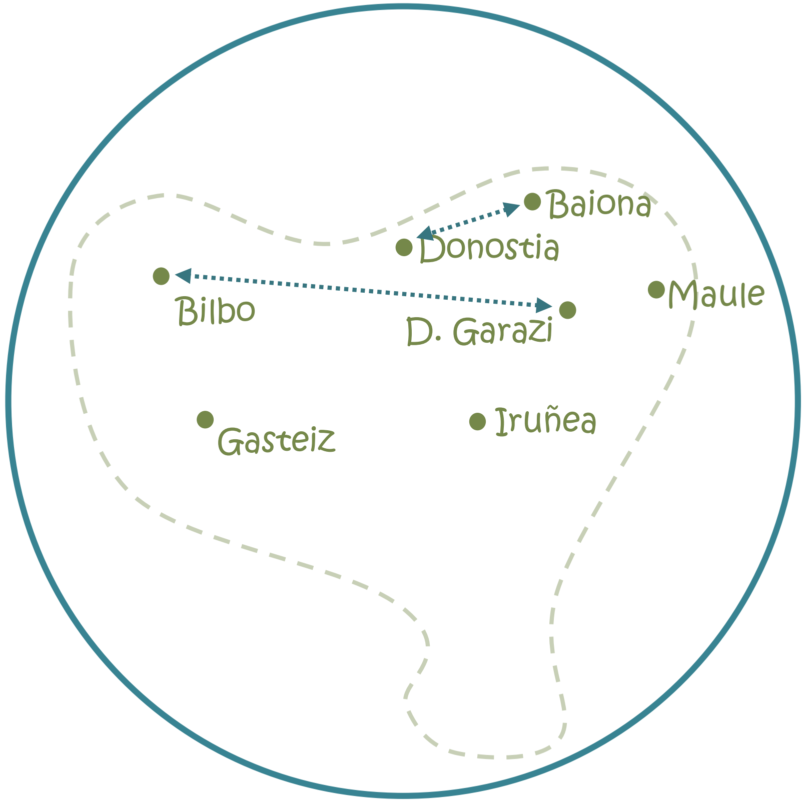
Embeddings



Geographical space

- Cities
- Meaningful distances

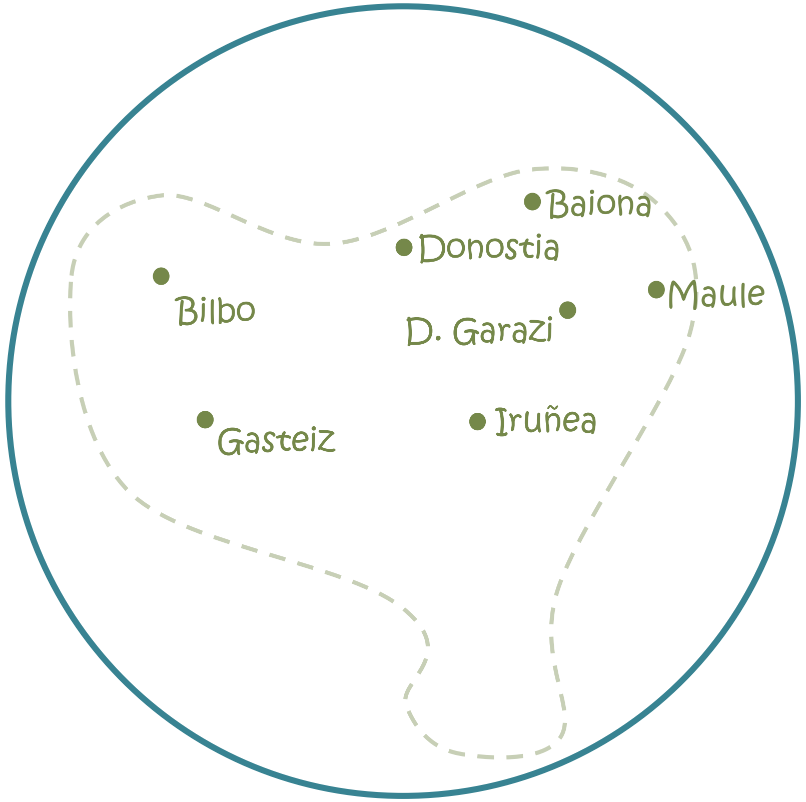
Embeddings



<u>Geographical space</u>

- Cities
- Meaningful distances

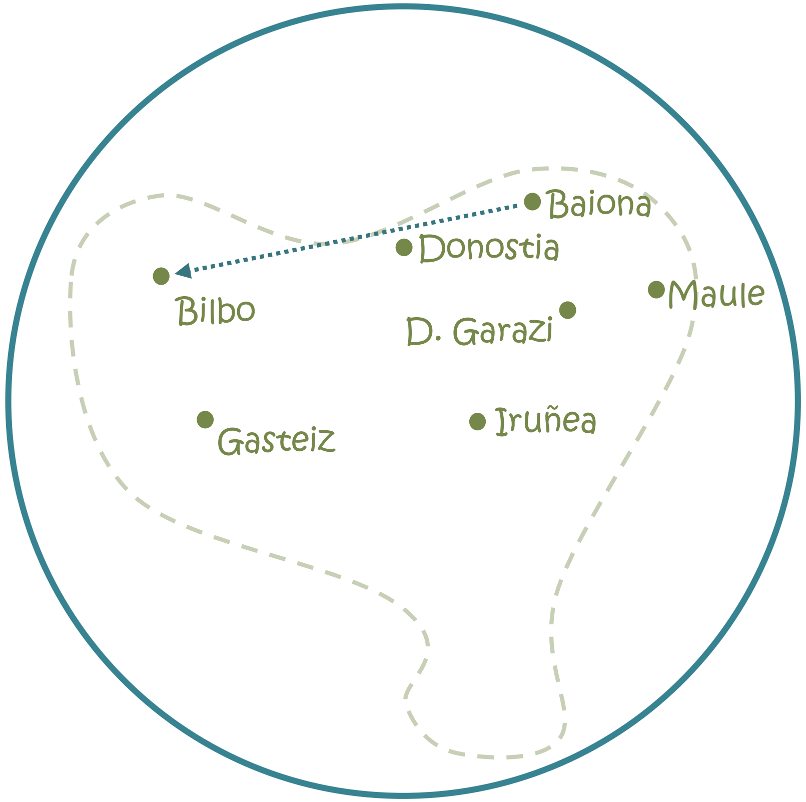
Embeddings



<u>Geographical space</u>

- Cities
- Meaningful distances
- Meaningful relations

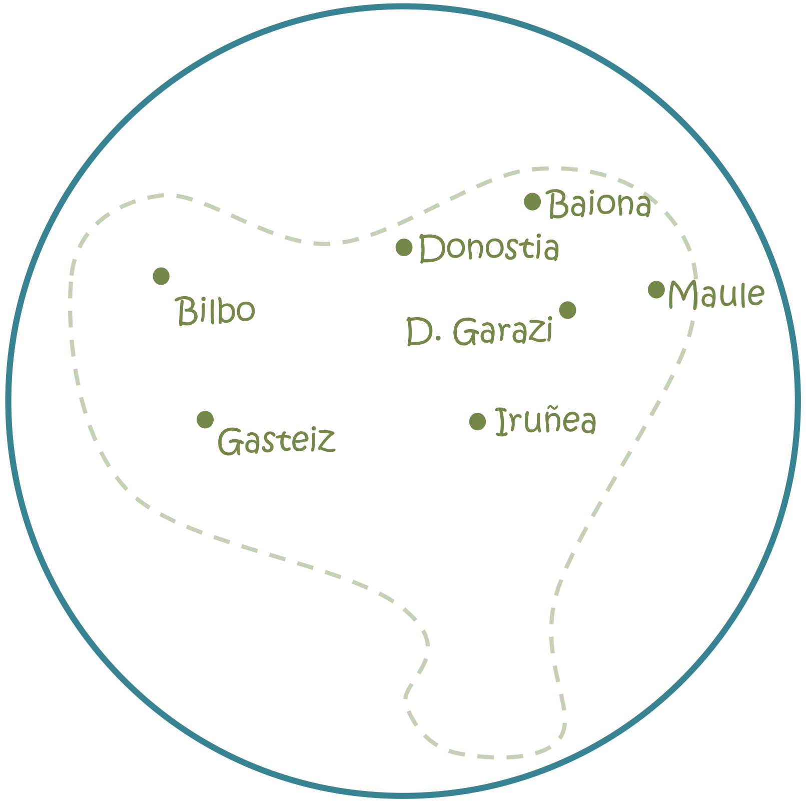
Embeddings



<u>Geographical space</u>

- Cities
- Meaningful distances
- Meaningful relations

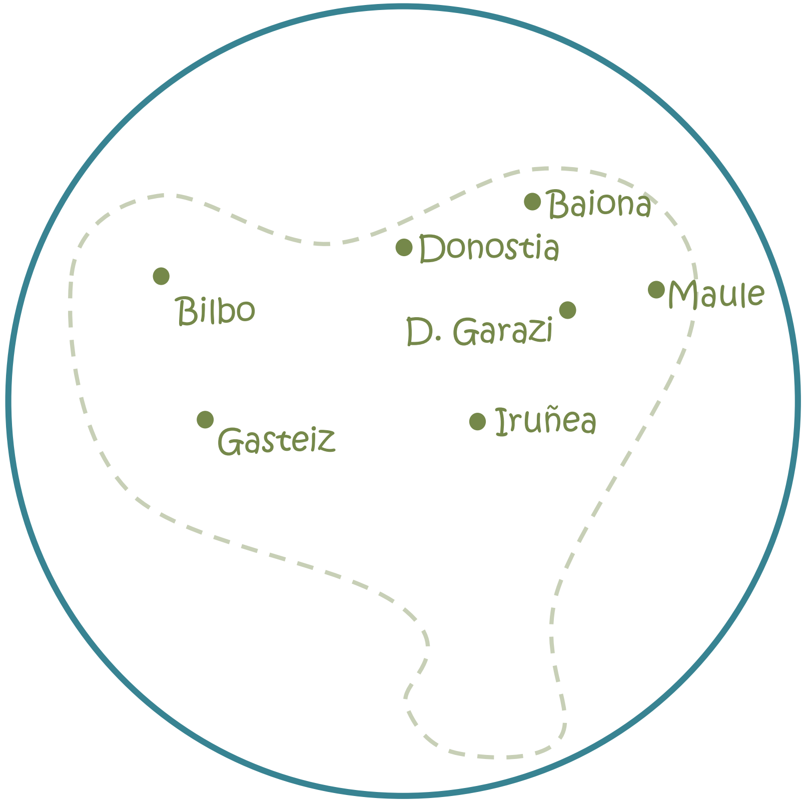
Embeddings



<u>Geographical space</u>

- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions

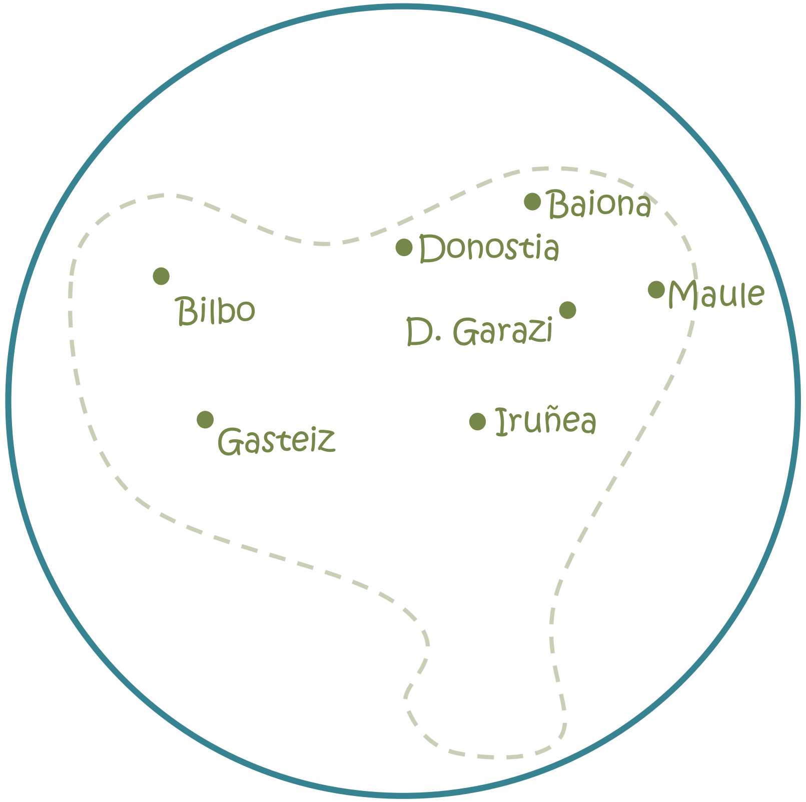
Embeddings



<u>Geographical space</u>

- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Embeddings

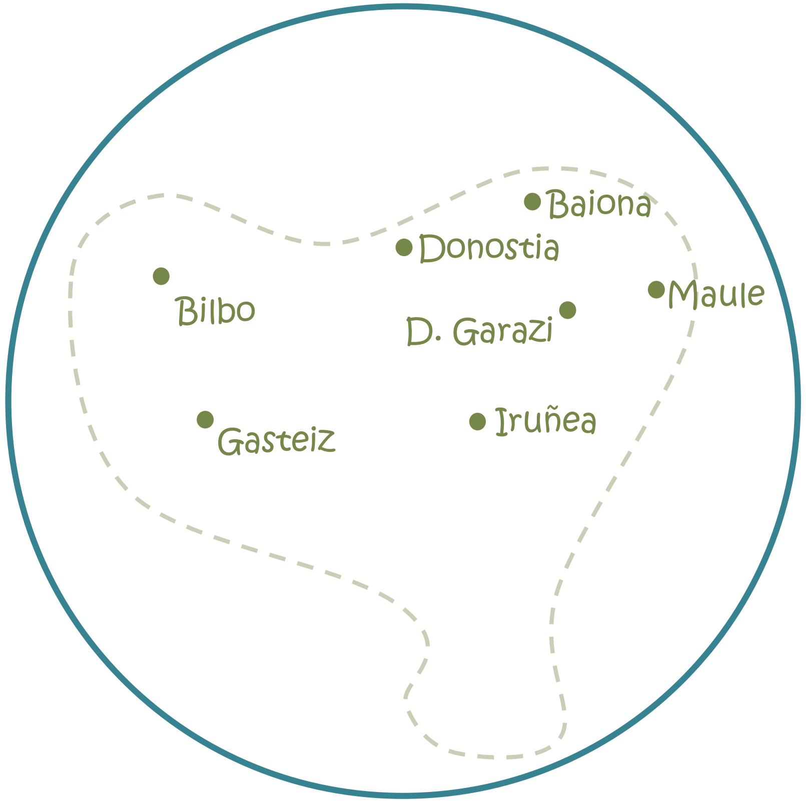


Geographical space

- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

Embeddings



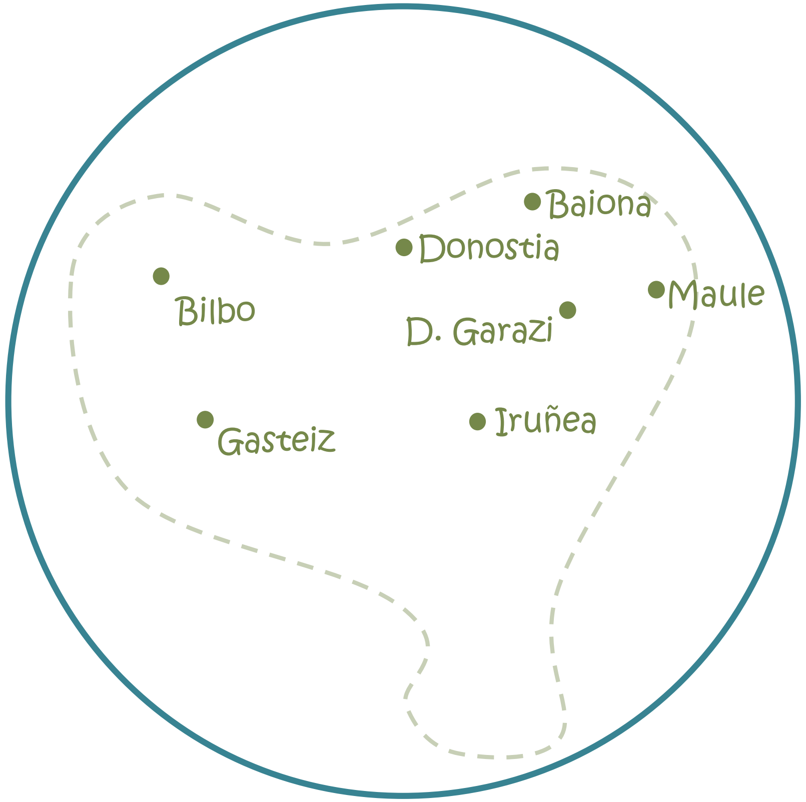
Geographical space

- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

- Words

Embeddings

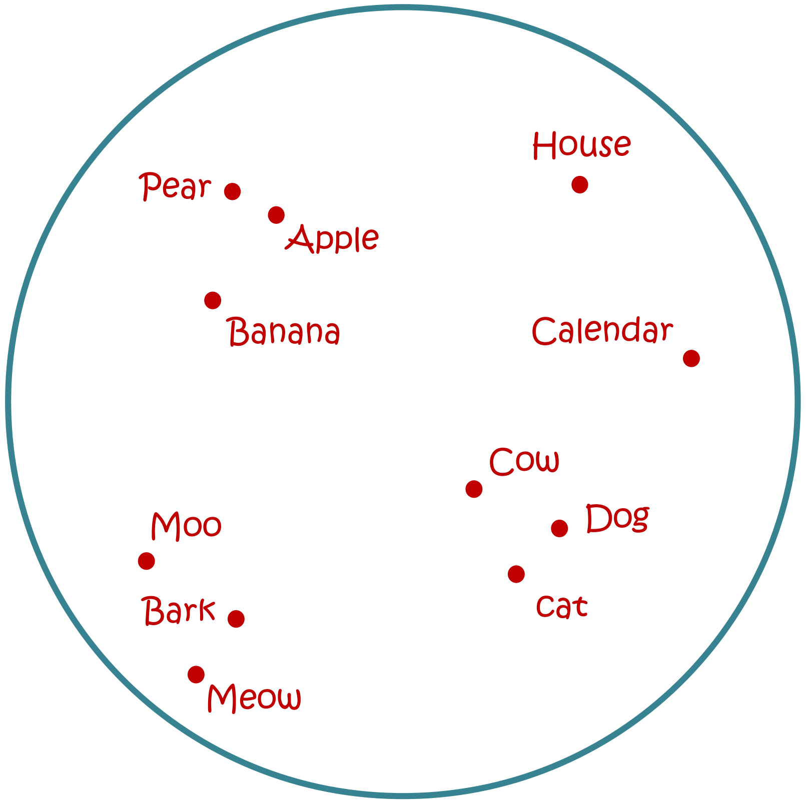


Geographical space

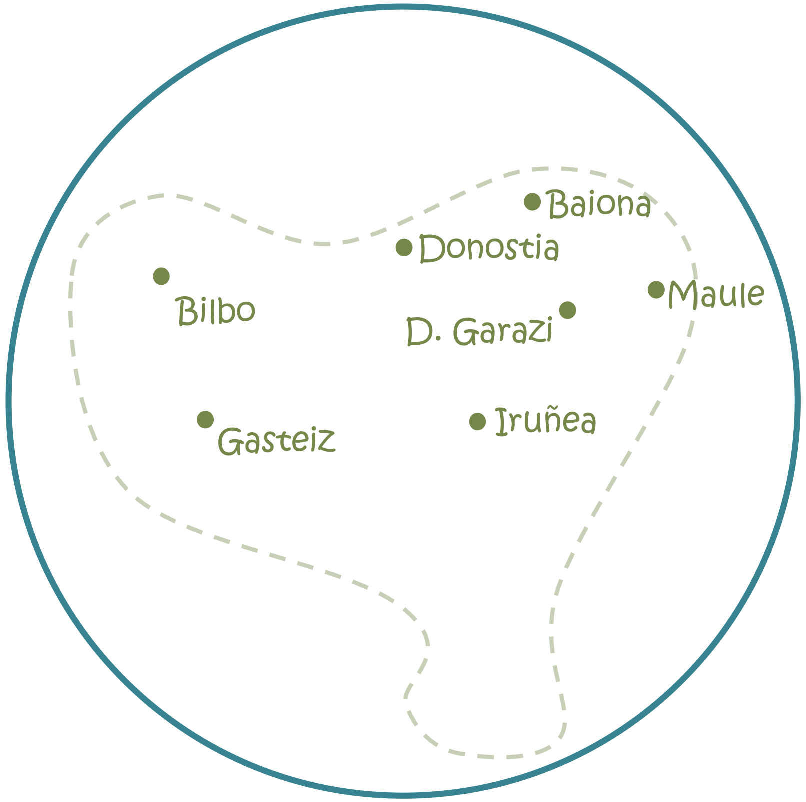
- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

- Words



Embeddings

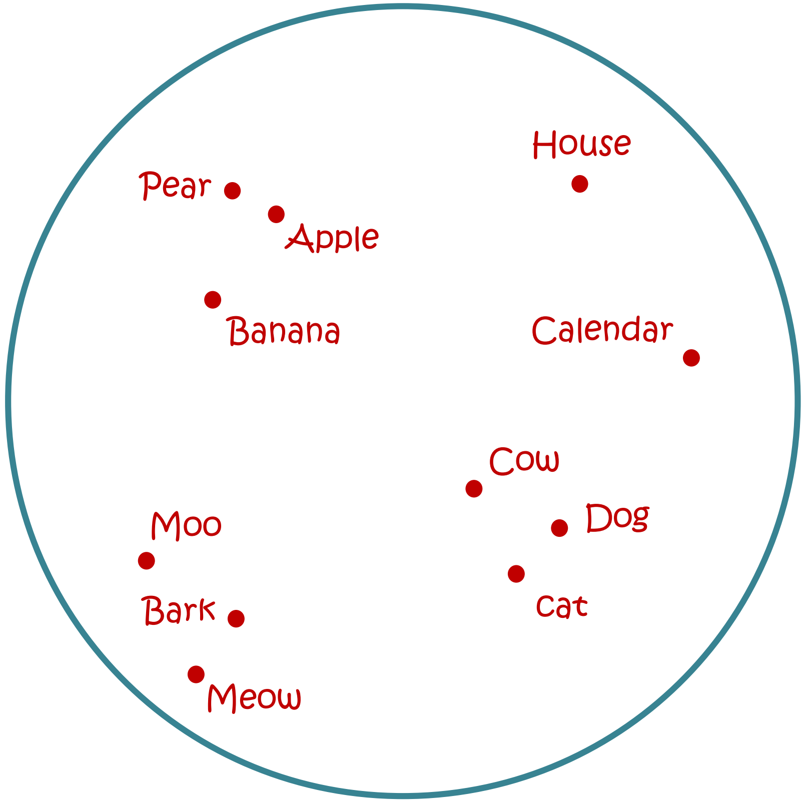


Geographical space

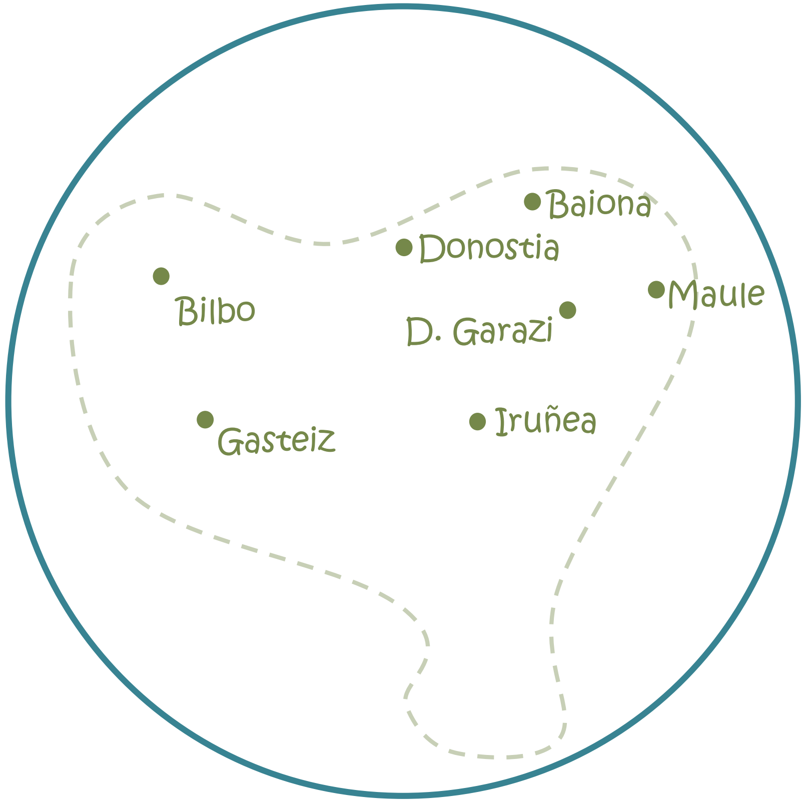
- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

- Words
- Meaningful distances



Embeddings

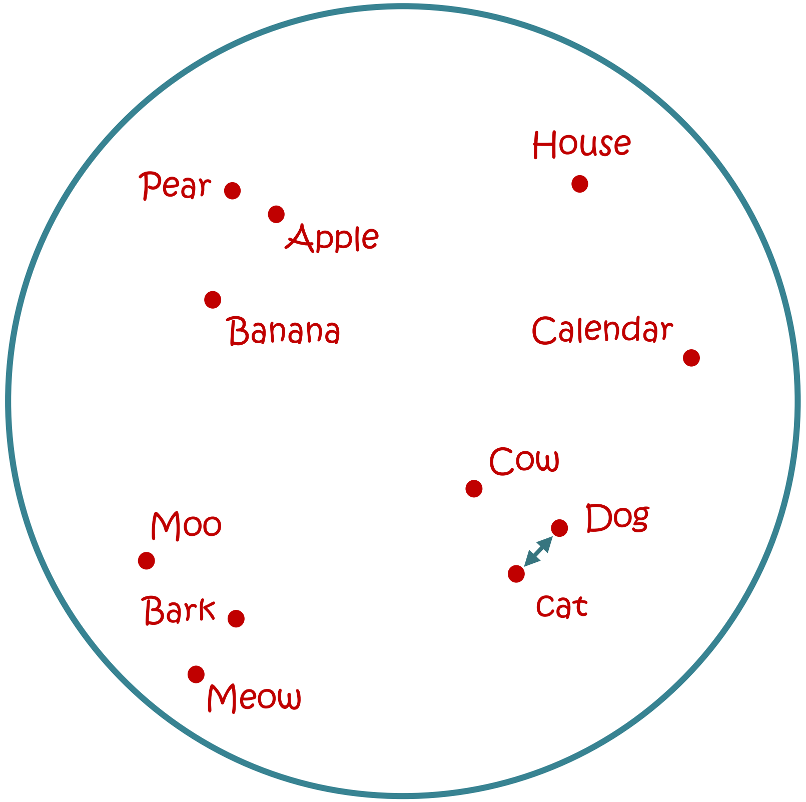


Geographical space

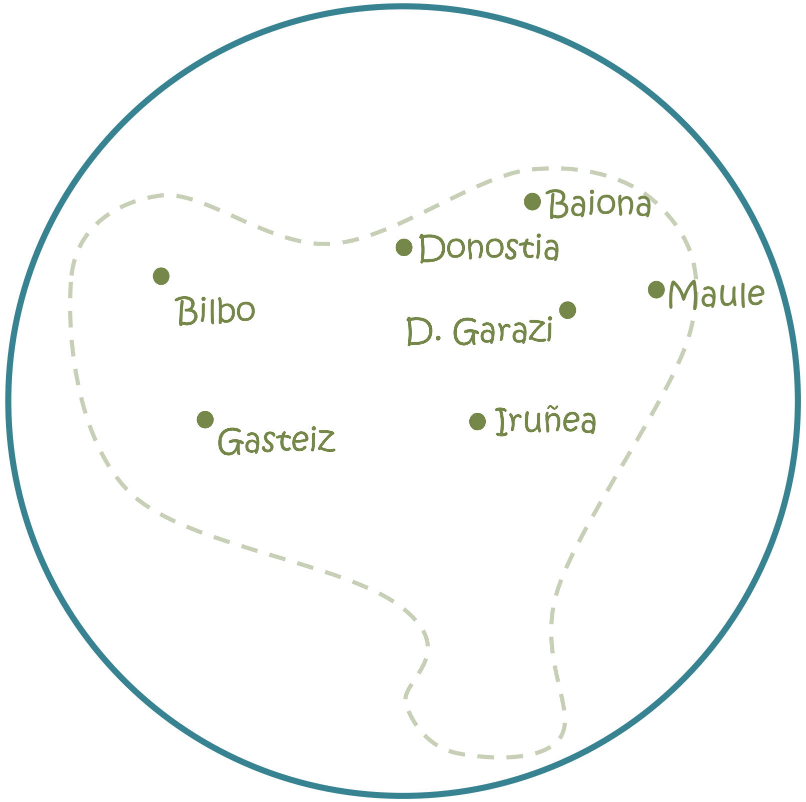
- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

- Words
- Meaningful distances



Embeddings

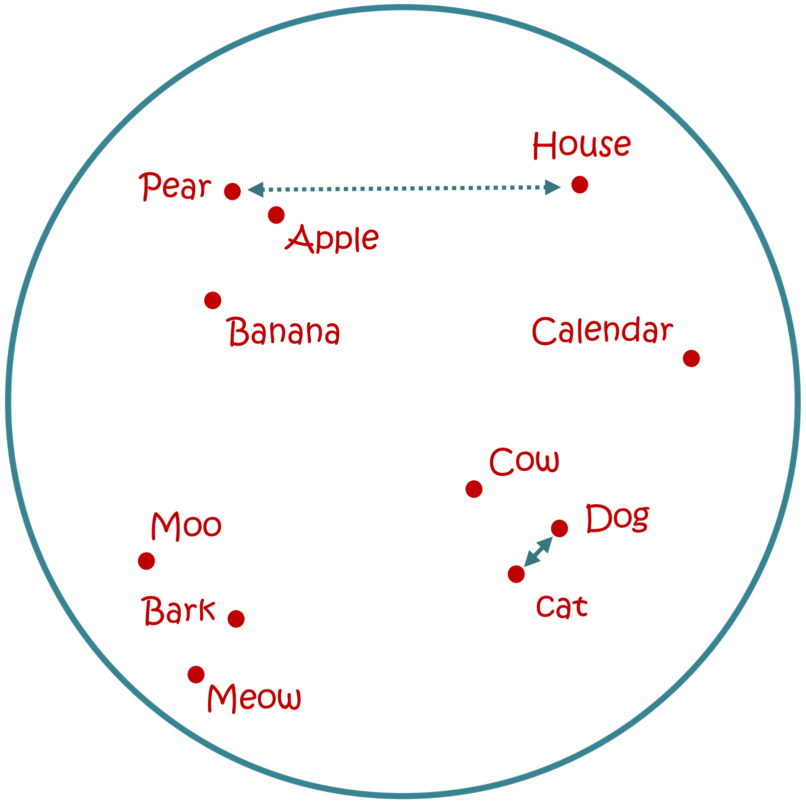


Geographical space

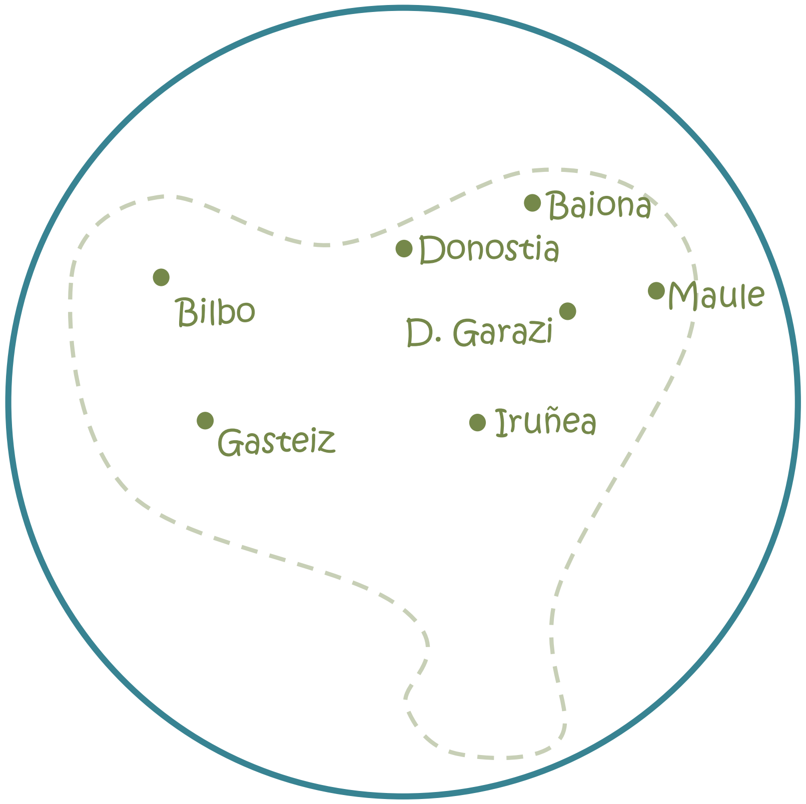
- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

- Words
- Meaningful distances



Embeddings

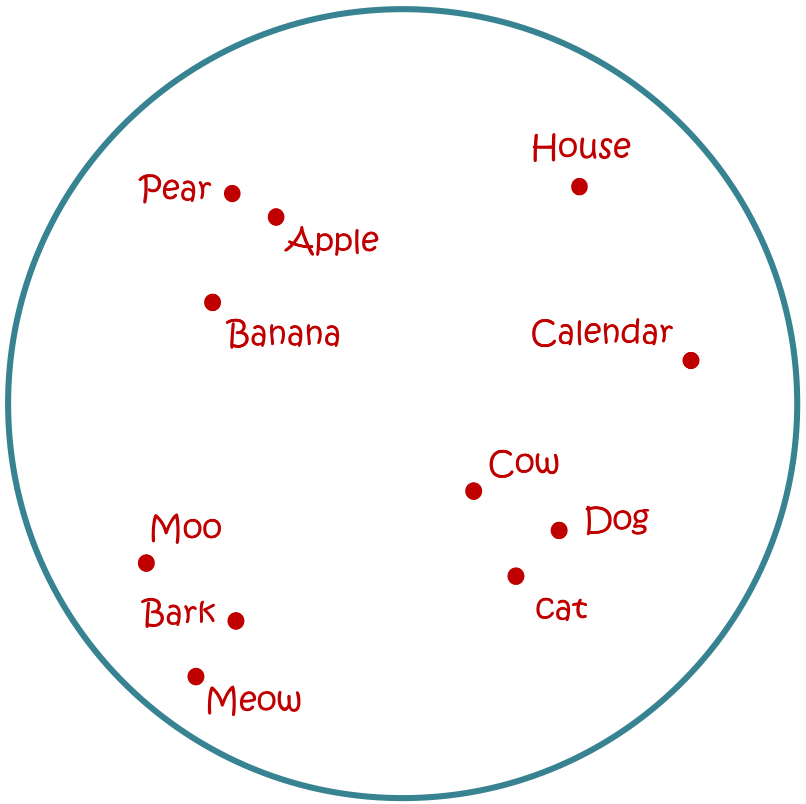


Geographical space

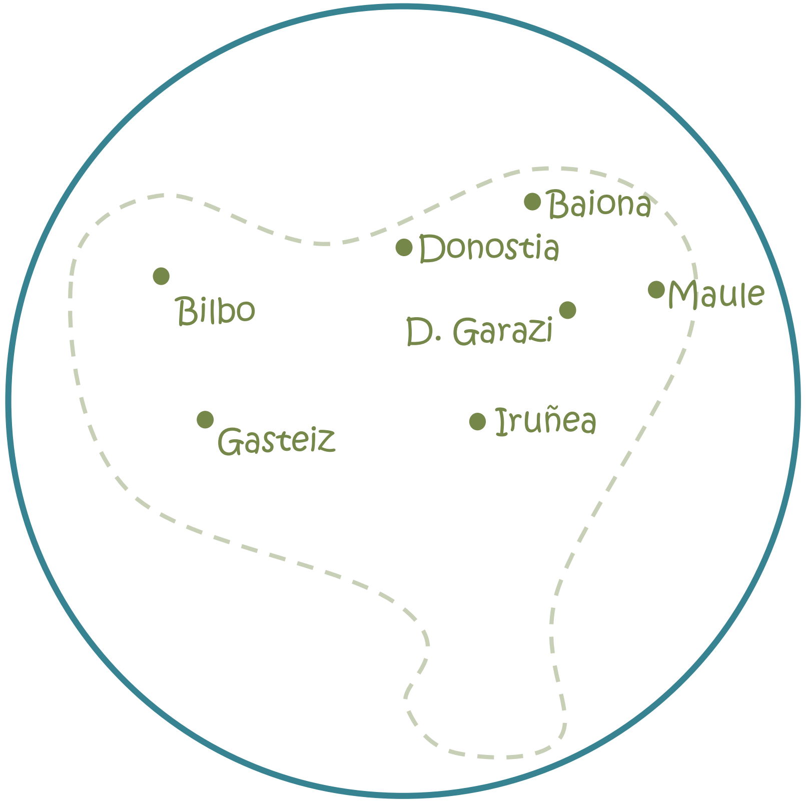
- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

- Words
- Meaningful distances
- Meaningful relations



Embeddings

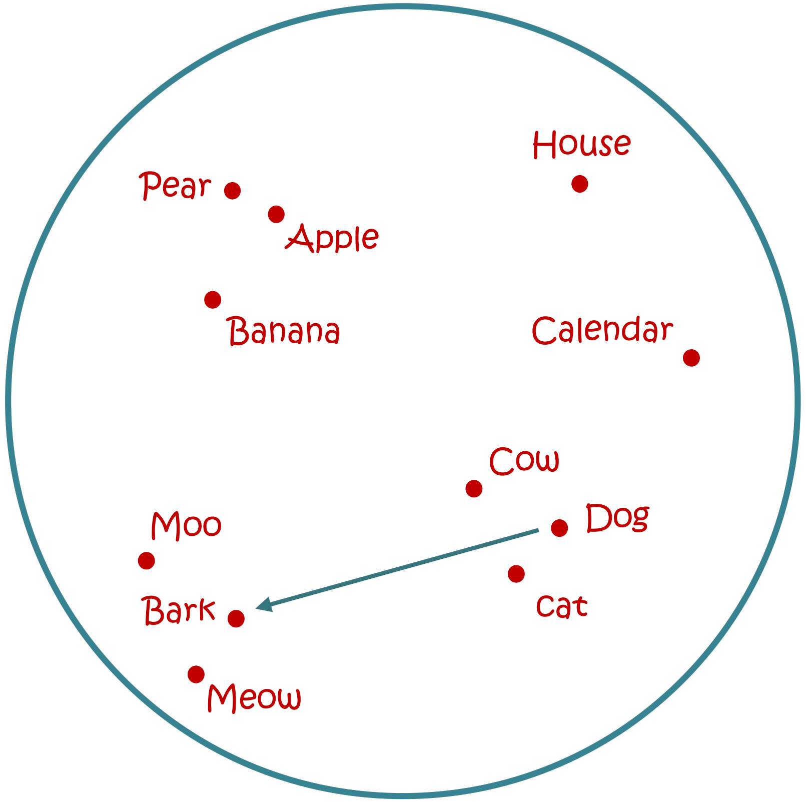


Geographical space

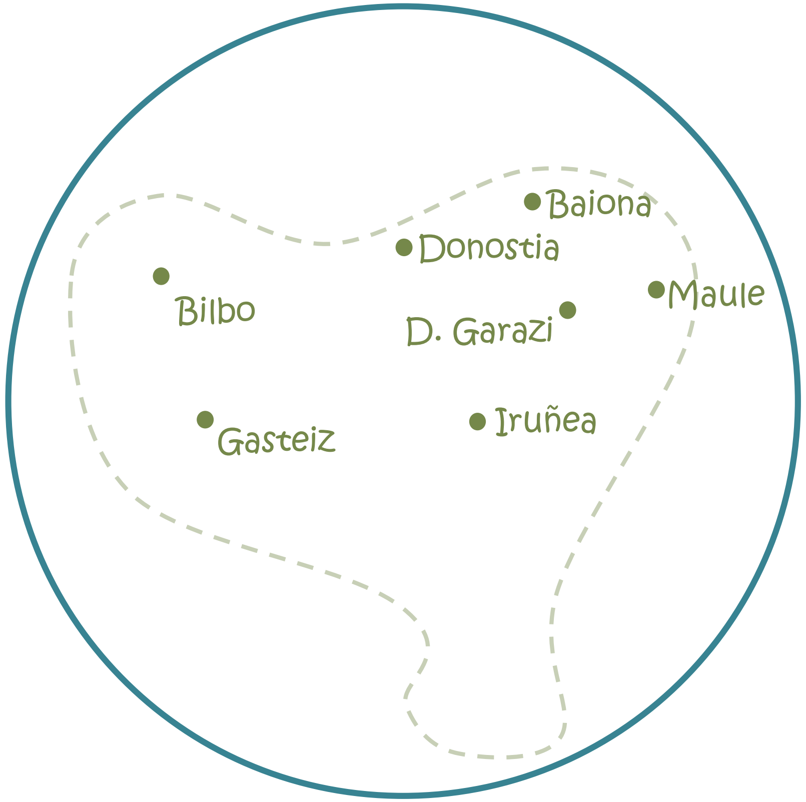
- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

- Words
- Meaningful distances
- Meaningful relations



Embeddings

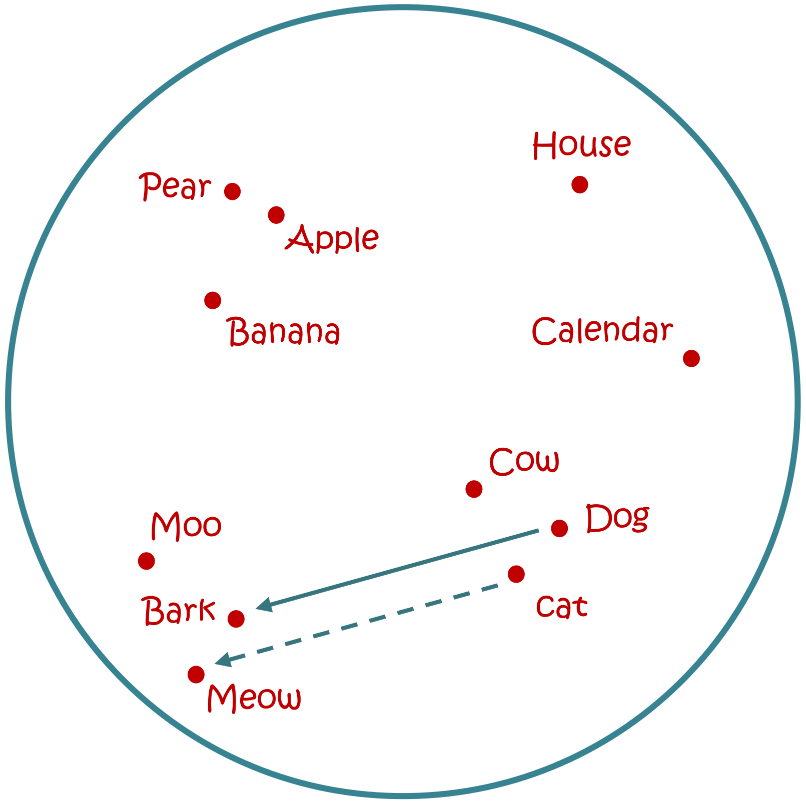


Geographical space

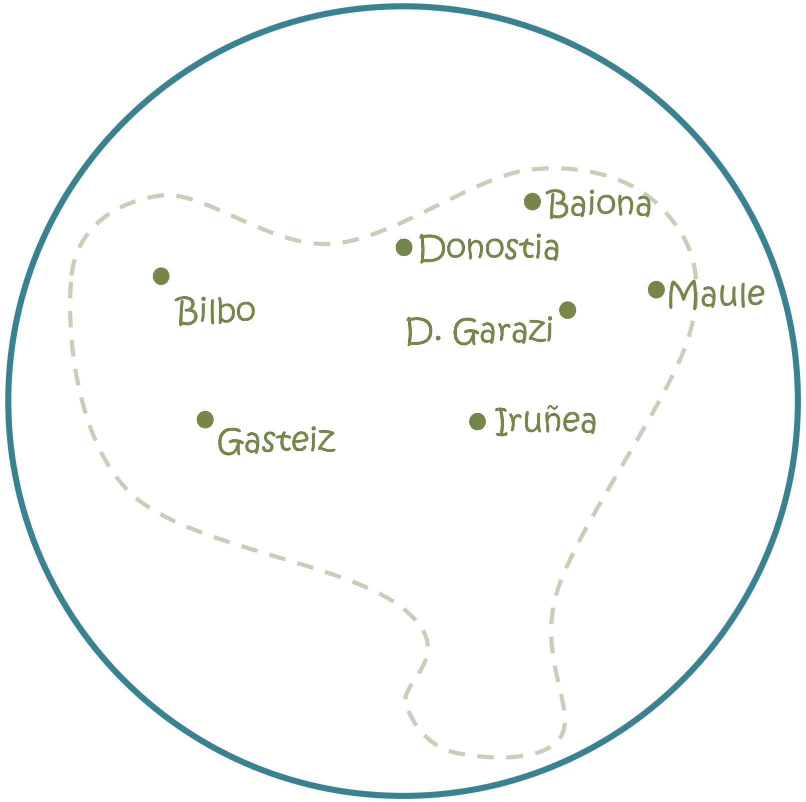
- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

- Words
- Meaningful distances
- Meaningful relations



Embeddings

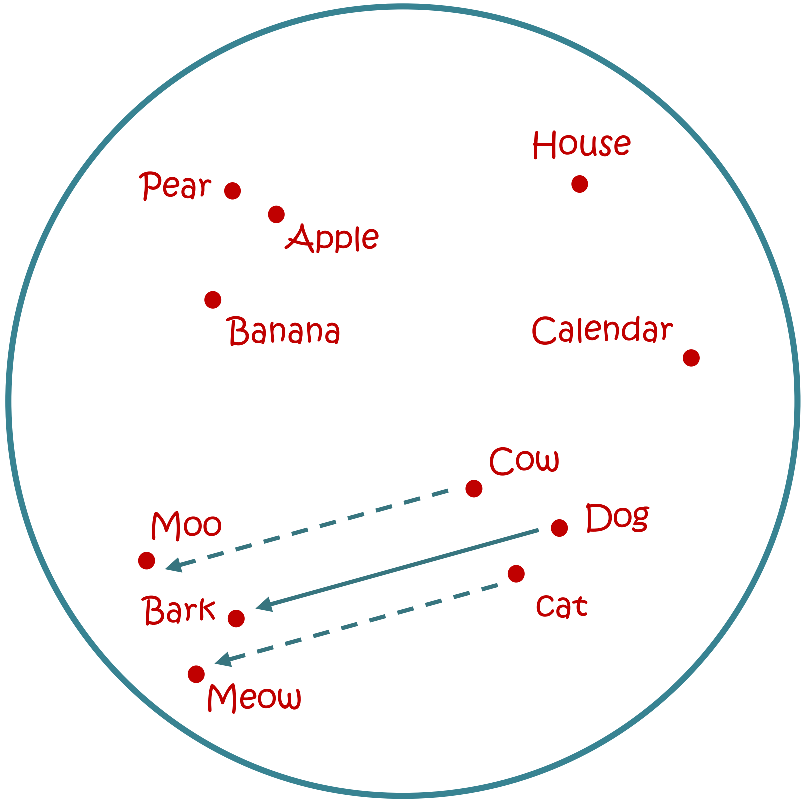


Geographical space

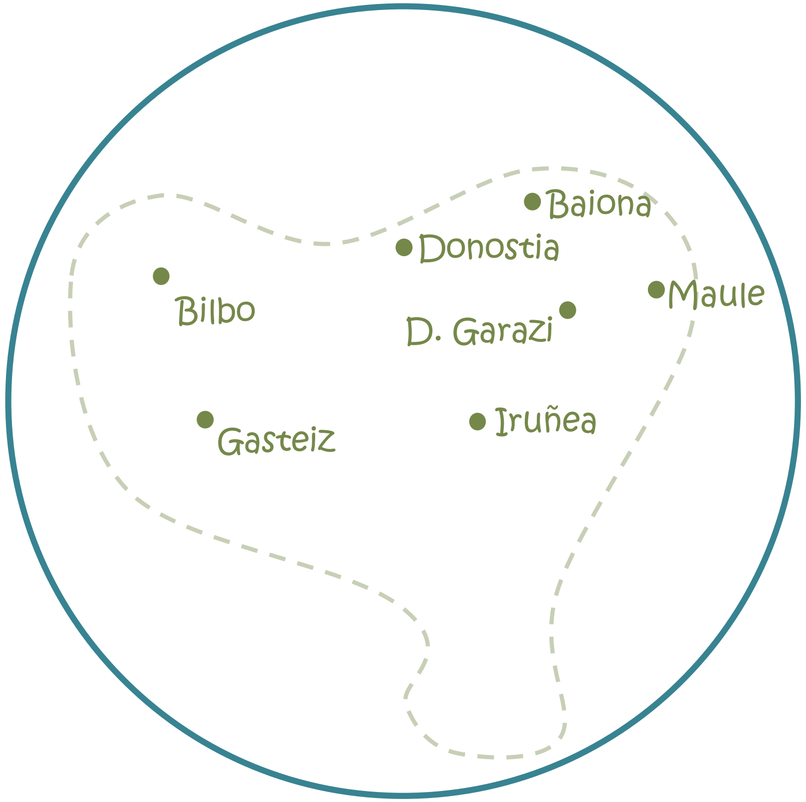
- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

- Words
- Meaningful distances
- Meaningful relations



Embeddings

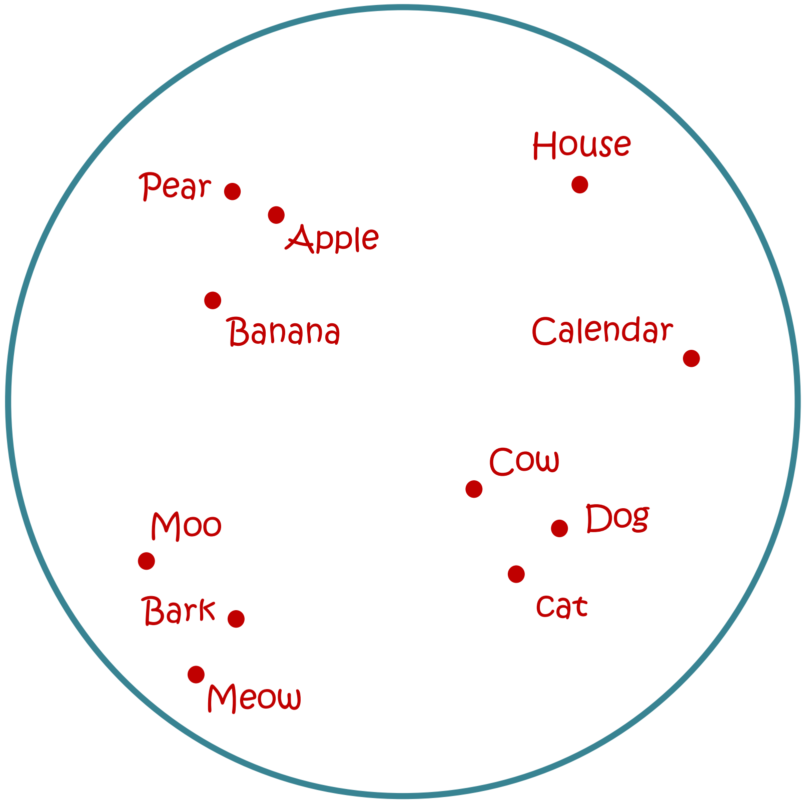


Geographical space

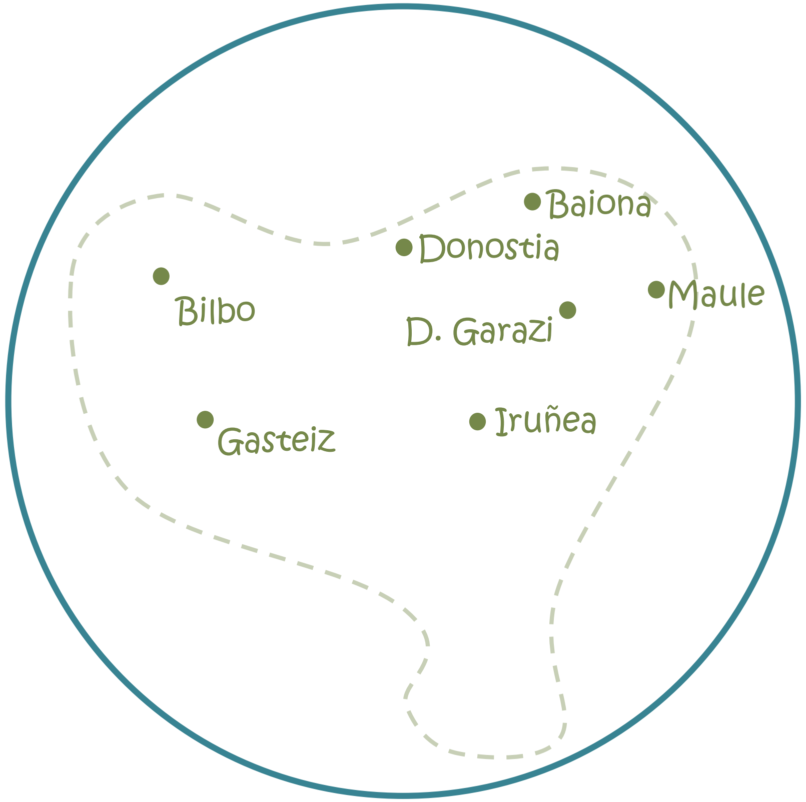
- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions



Embeddings

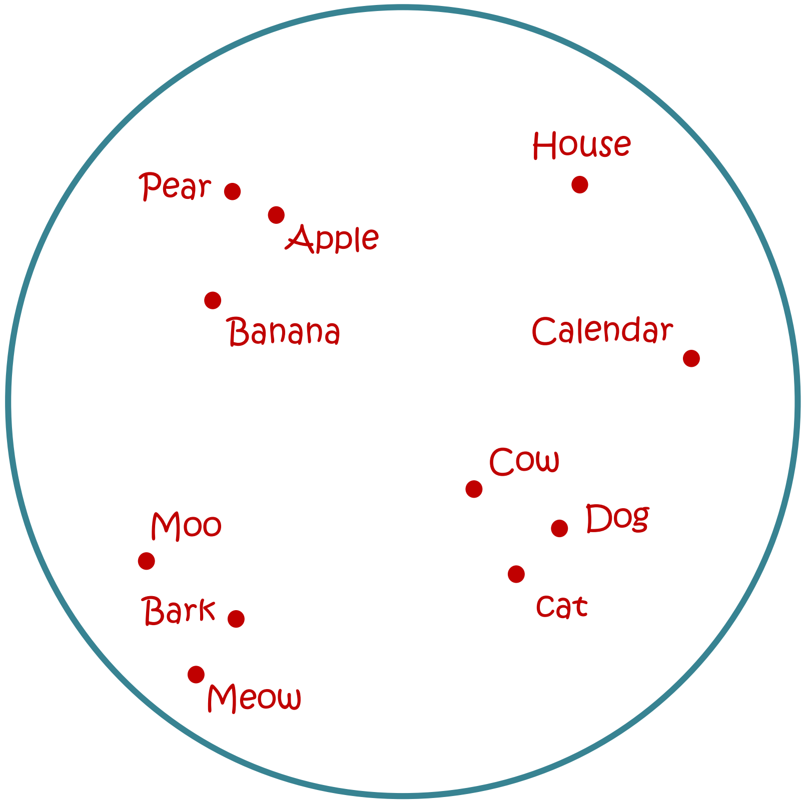


Geographical space

- Cities
- Meaningful distances
- Meaningful relations
- 2 dimensions
- Cartographers from 3D world

Semantic space

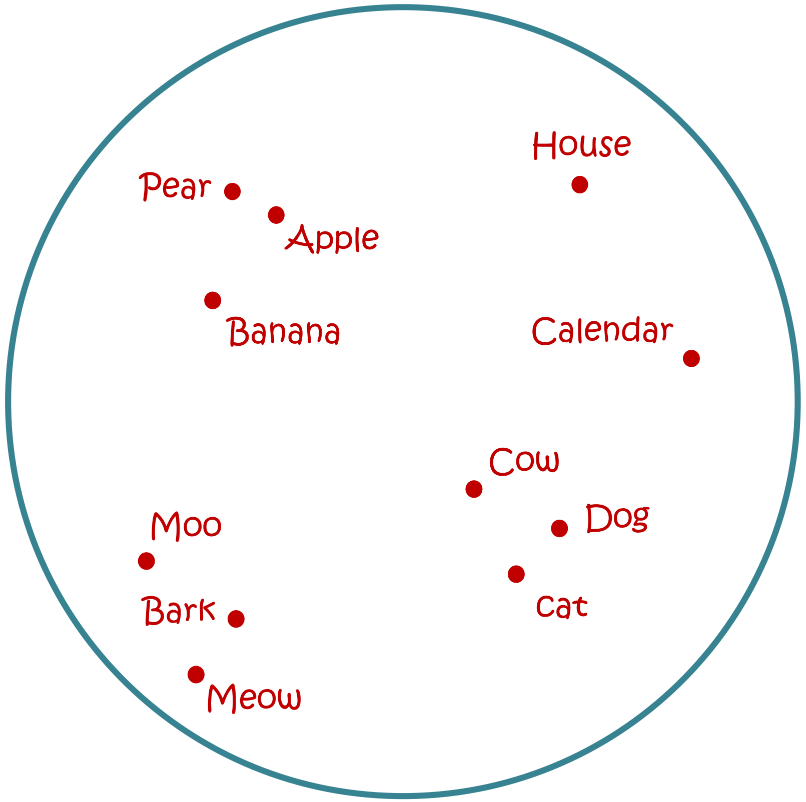
- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts

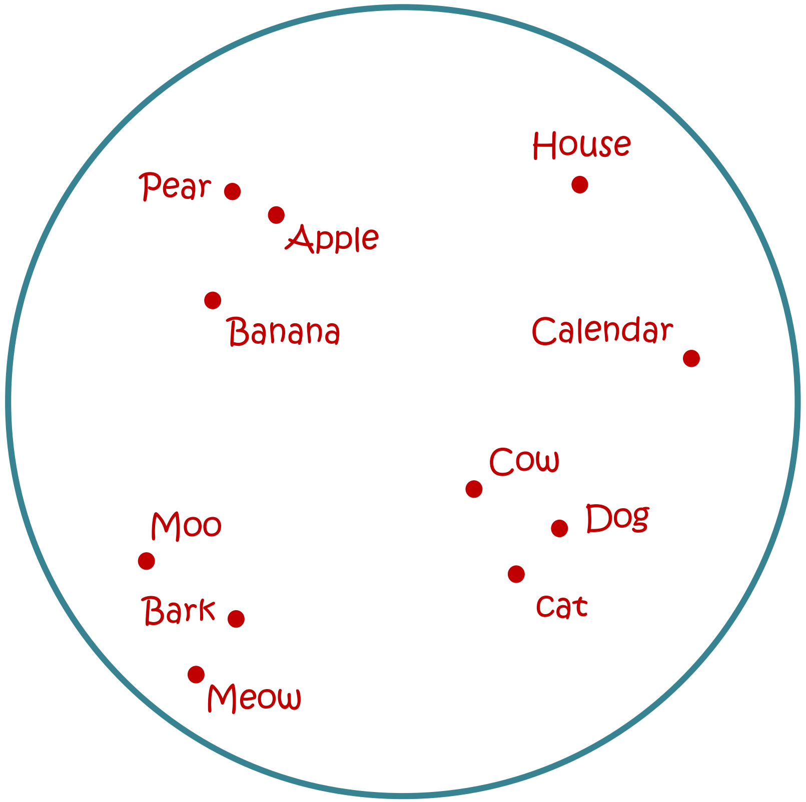


Embeddings

Traditional vector space models:

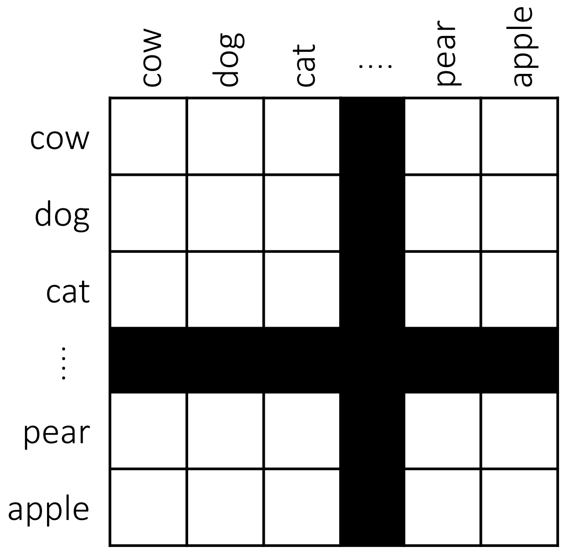
Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



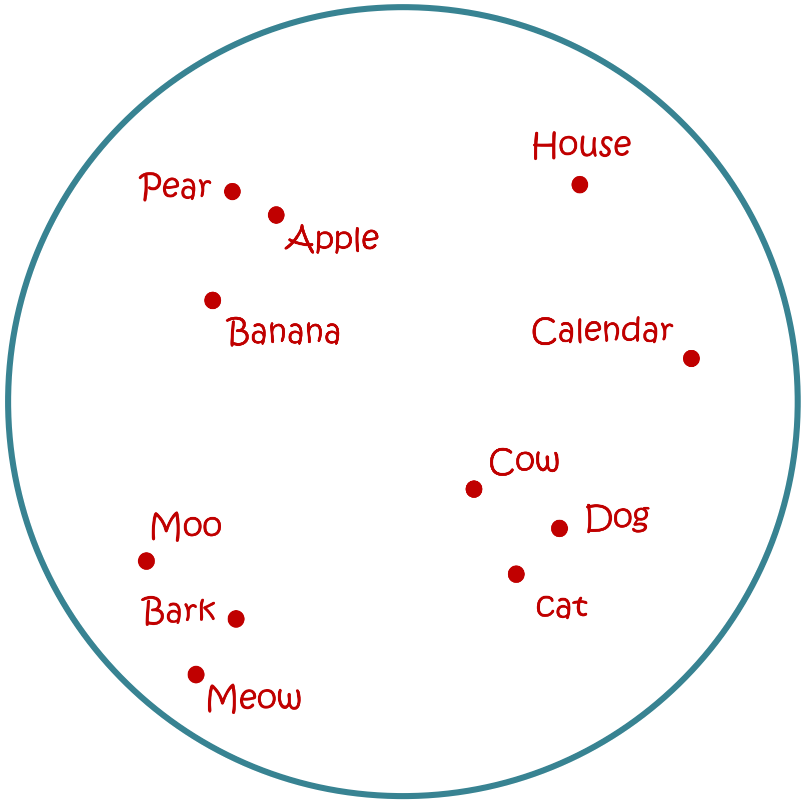
Embeddings

Traditional vector space models:



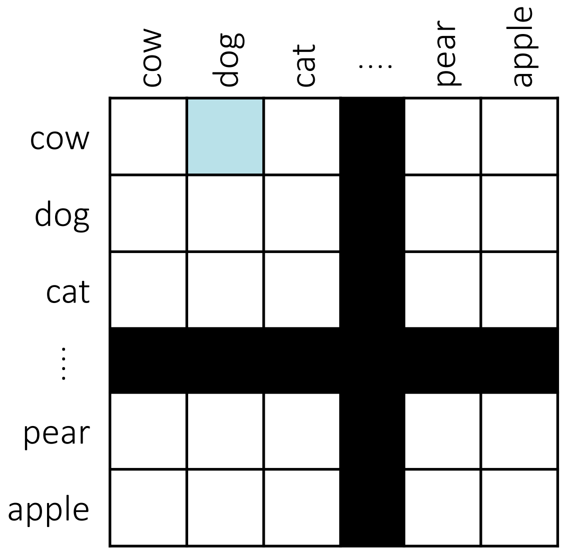
Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



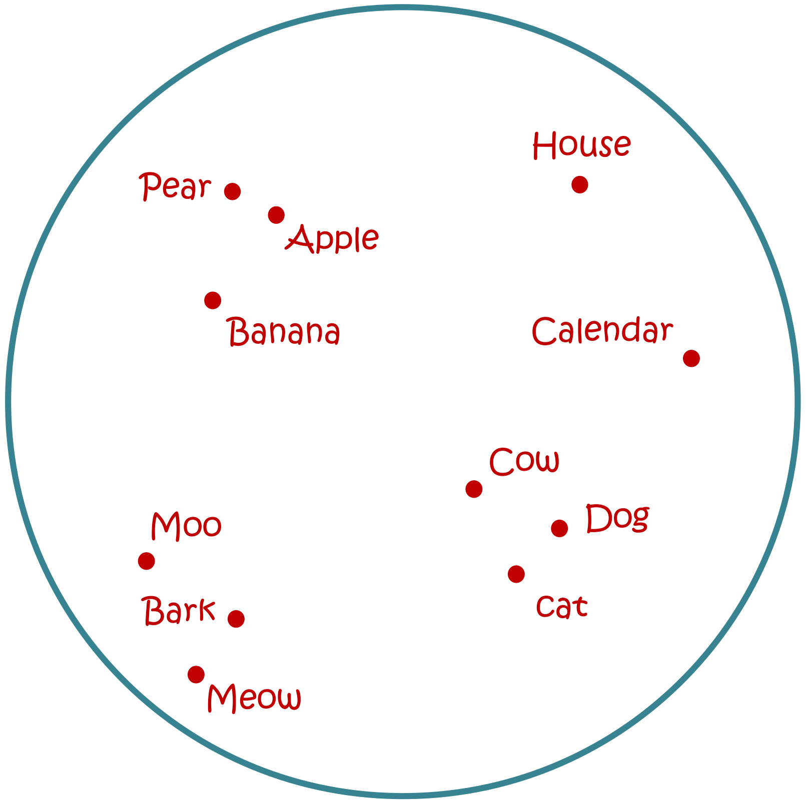
Embeddings

Traditional vector space models:



Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



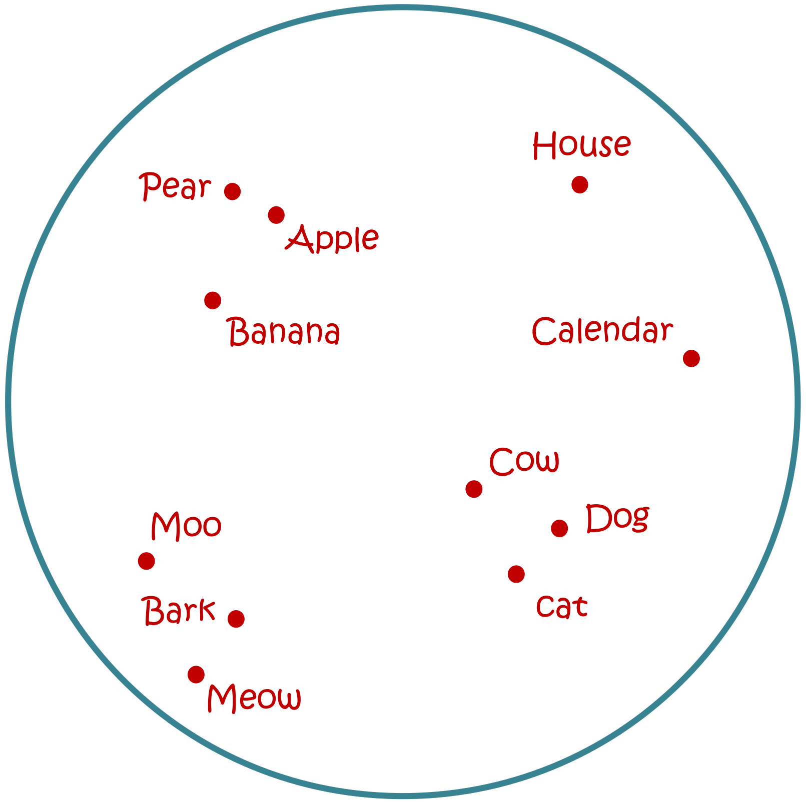
Embeddings

Traditional vector space models:

	cow	dog	cat	....	pear	apple
cow		7				
dog						
cat						
⋮						
pear						
apple						

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



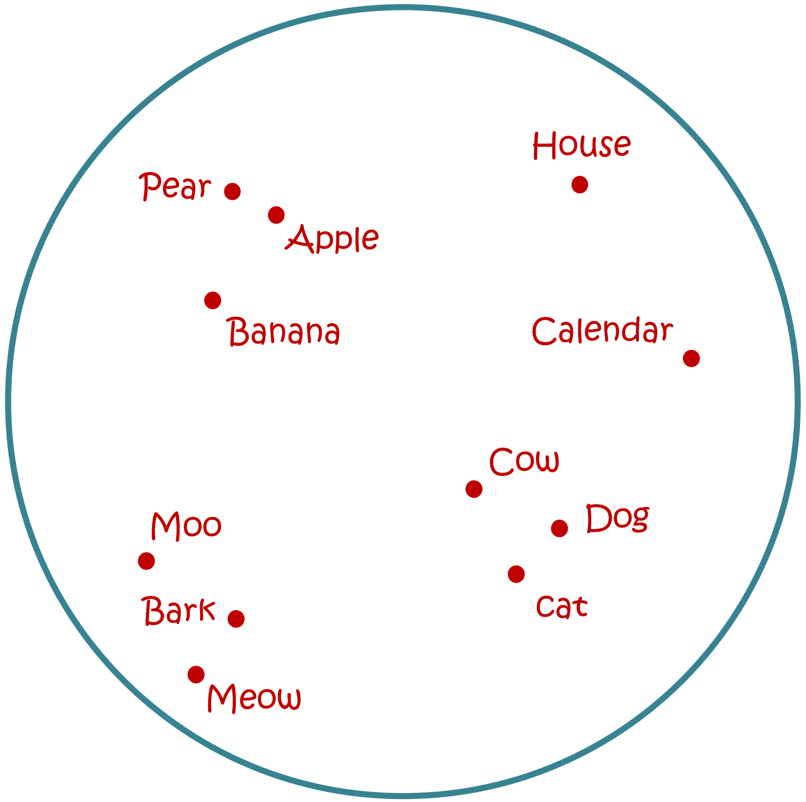
Embeddings

Traditional vector space models:

	cow	dog	cat	....	pear	apple
cow		7				
dog						
cat						
⋮						
pear						
apple						

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



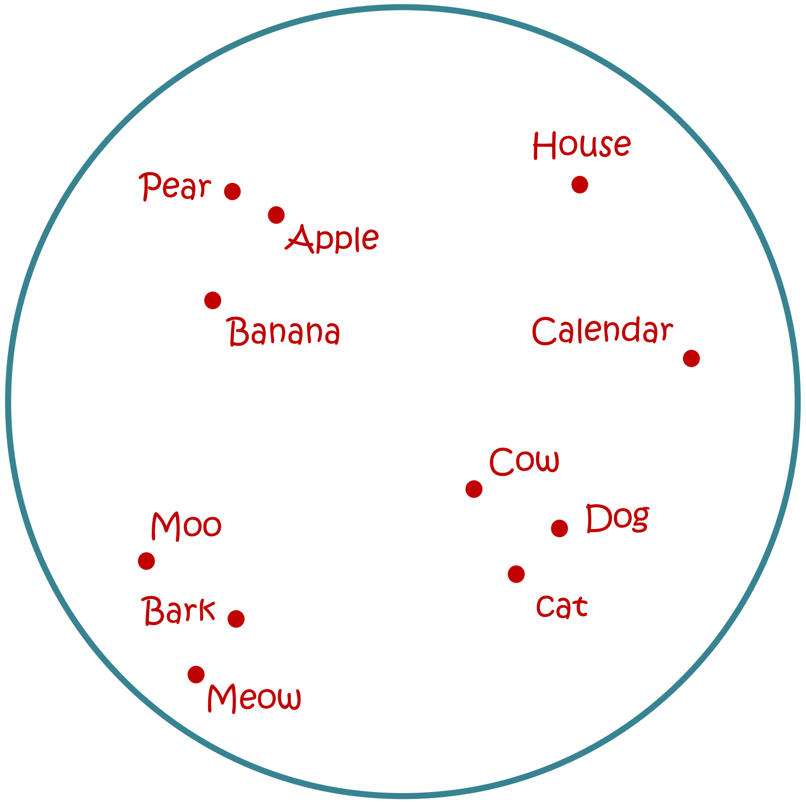
Embeddings

Traditional vector space models:

	cow	dog	cat	....	pear	apple
cow		7				
dog			24			
cat						
⋮						
pear						
apple						

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



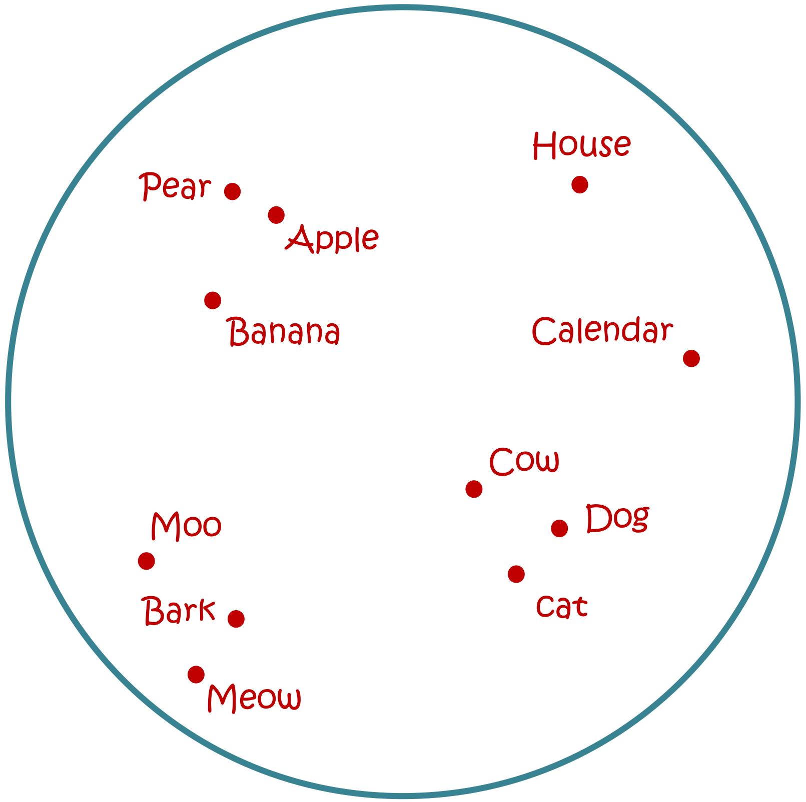
Embeddings

Traditional vector space models:

	cow	dog	cat	....	pear	apple
cow	23	7	4		1	2
dog	7	34	24		2	3
cat	4	24	27		2	1
⋮						
pear	1	2	2		19	21
apple	2	3	1		21	28

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



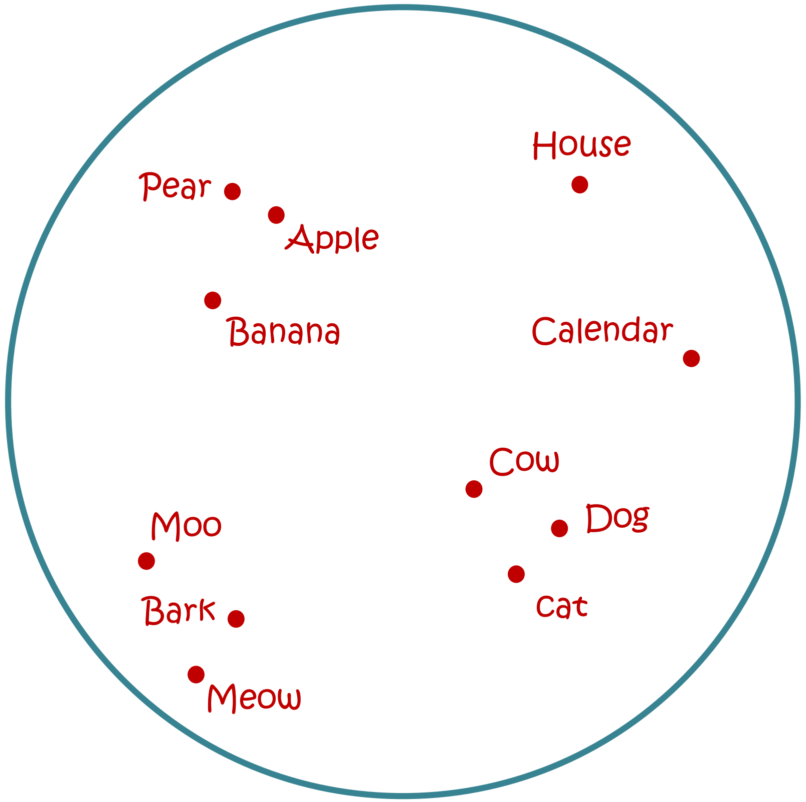
Embeddings

Traditional vector space models:

cow	23	7	4		1	2
dog	7	34	24		2	3
cat	4	24	27		2	1
⋮						
pear	1	2	2		19	21
apple	2	3	1		21	28

<u>Semantic space</u>

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



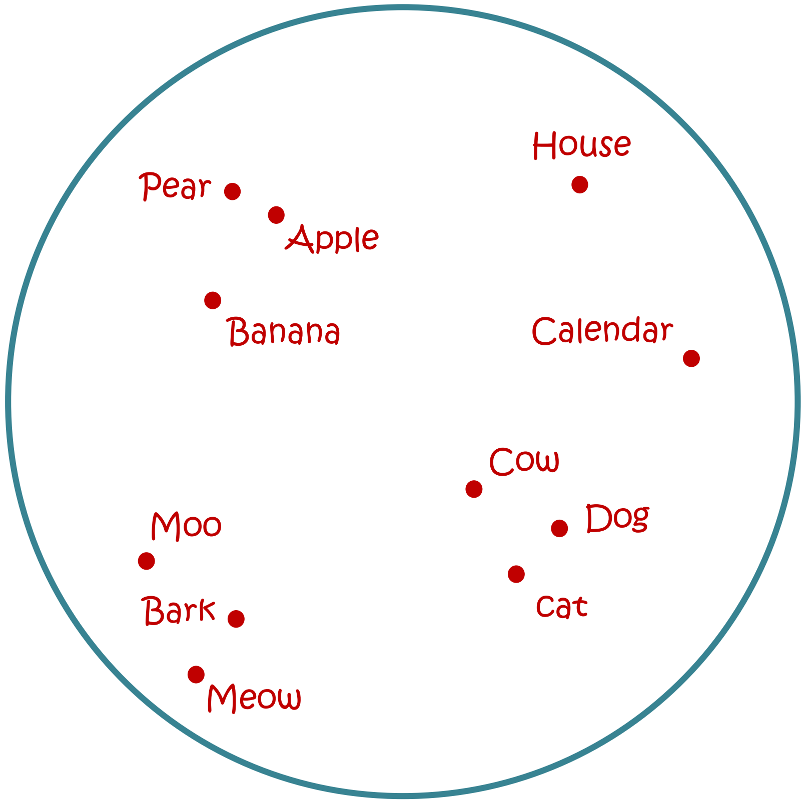
Embeddings

Traditional vector space models:

cow	23	7	4		1	2
dog	7	34	24		2	3
cat	4	24	27		2	1
⋮						
pear	1	2	2		19	21
apple	2	3	1		21	28

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

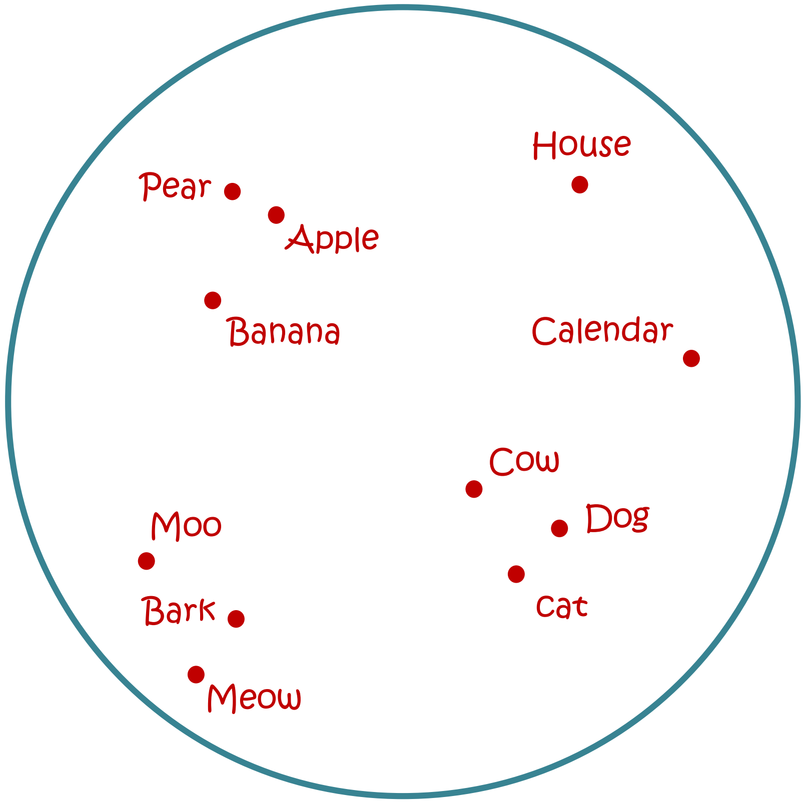
Traditional vector space models:

cow	23	7	4		1	2
dog	7	34	24		2	3
cat	4	24	27		2	1
⋮						
pear	1	2	2		19	21
apple	2	3	1		21	28

cow: (23, 7, 4, …, 1, 2)
dog: (7, 34, 24, …, 2, 3)
cat: (4, 24, 27, …, 2, 1)
⋮
pear: (1, 2, 2, …, 19, 21)
apple: (2, 3, 1, …, 21, 28)

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

Traditional vector space models:

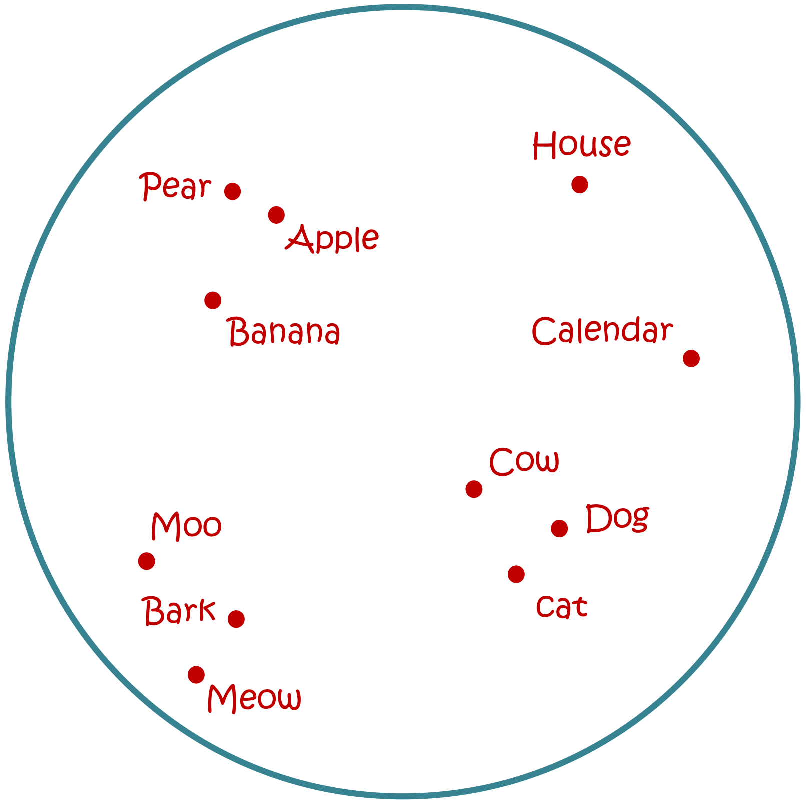
cow	23	7	4		1	2
dog	7	34	24		2	3
cat	4	24	27		2	1
⋮						
pear	1	2	2		19	21
apple	2	3	1		21	28

cow: (23, 7, 4, …, 1, 2)
dog: (7, 34, 24, …, 2, 3)
cat: (4, 24, 27, …, 2, 1)
⋮
pear: (1, 2, 2, …, 19, 21)
apple: (2, 3, 1, …, 21, 28)

PCA

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

Traditional vector space models:

cow	23	7	4		1	2
dog	7	34	24		2	3
cat	4	24	27		2	1
⋮						
pear	1	2	2		19	21
apple	2	3	1		21	28

cow: (23, 7, 4, ..., 1, 2)

dog: (7, 34, 24, ..., 2, 3)

cat: (4, 24, 27, ..., 2, 1)

⋮

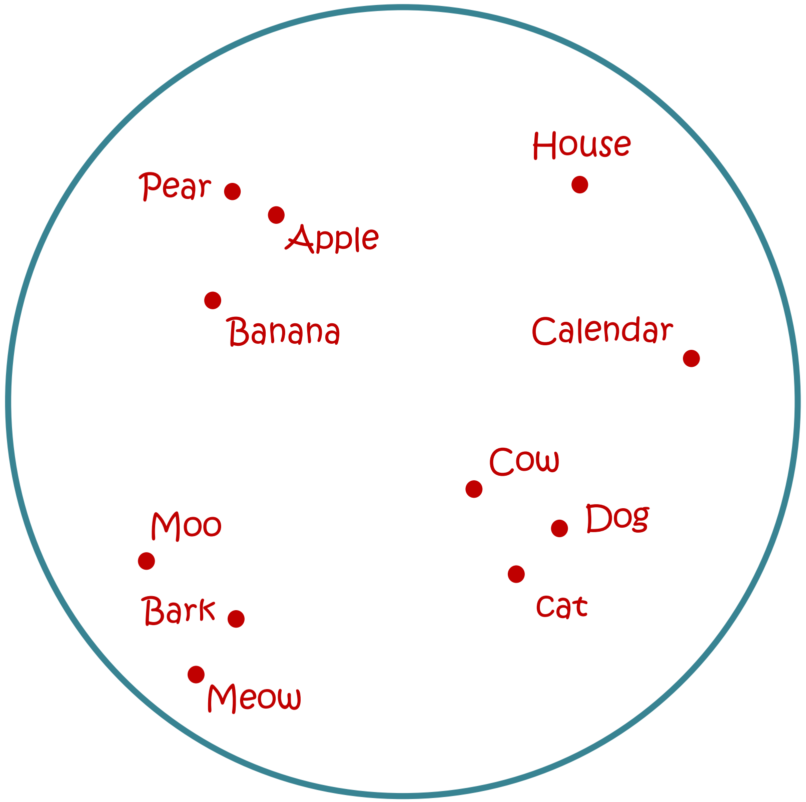
pear: (1, 2, 2, ..., 19, 21)

apple: (2, 3, 1, ..., 21, 28)

PCA

Semantic space

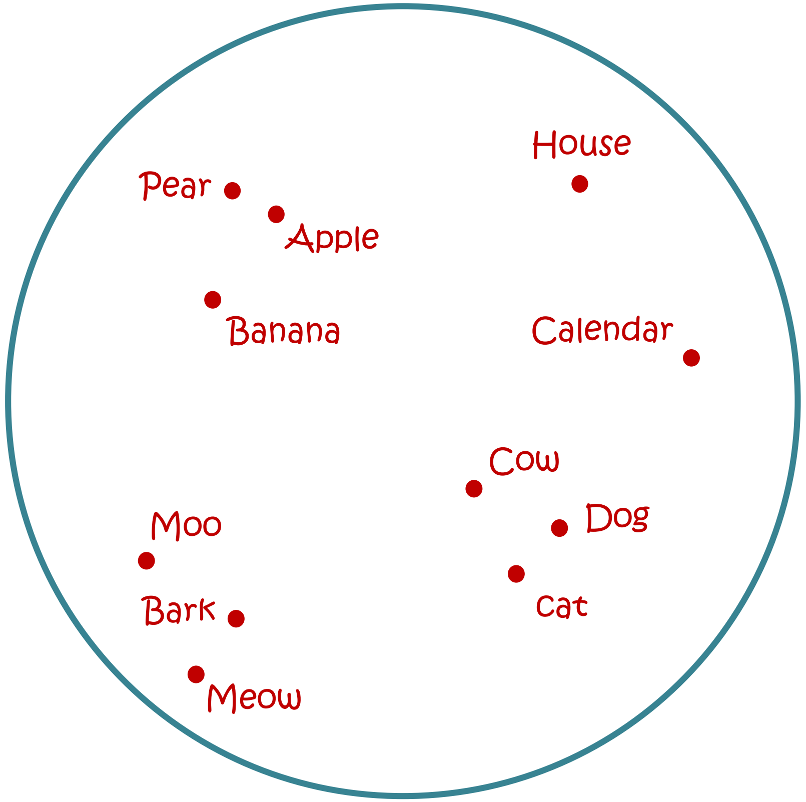
- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts

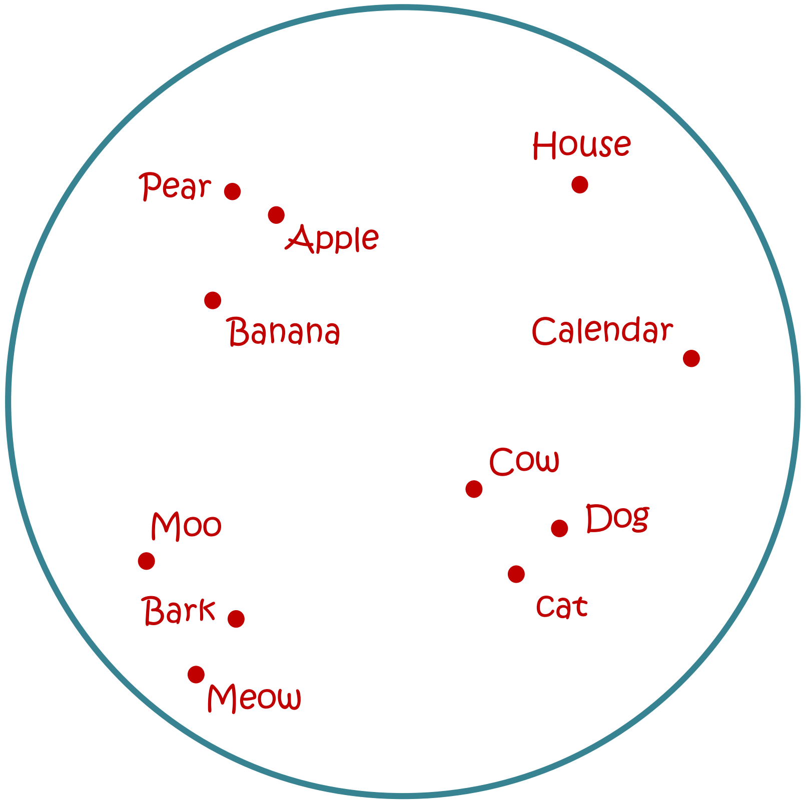


Embeddings

Skip-gram with negative sampling:

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



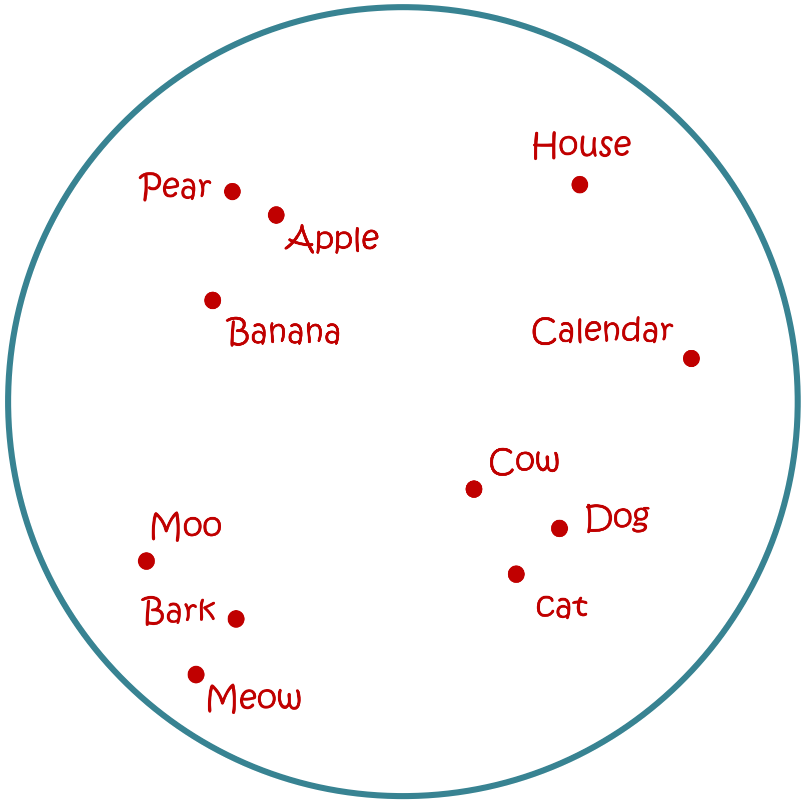
Embeddings

Skip-gram with negative sampling:

$$\log \sigma(w \cdot c) + \sum_{i=1}^{k} \mathbb{E}_{c_N \sim P_D}[\log \sigma(-w \cdot c_N)]$$

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

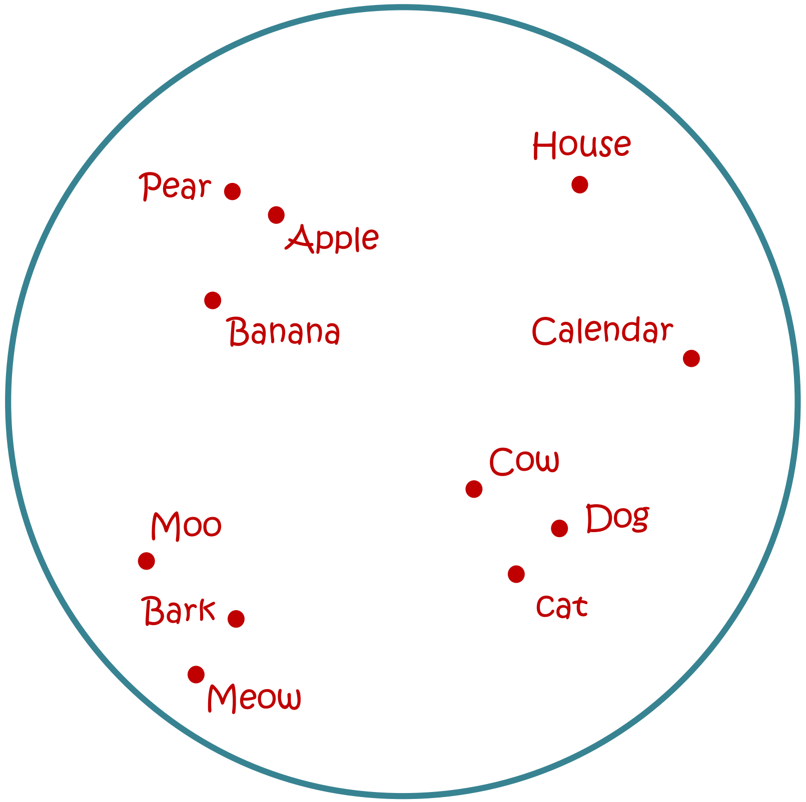
Skip-gram with negative sampling:

$$\log \sigma(w \cdot c) + \sum_{i=1}^{k} \mathbb{E}_{c_N \sim P_D}[\log \sigma(-w \cdot c_N)]$$

I will go to New York by plane .

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

Skip-gram with negative sampling:

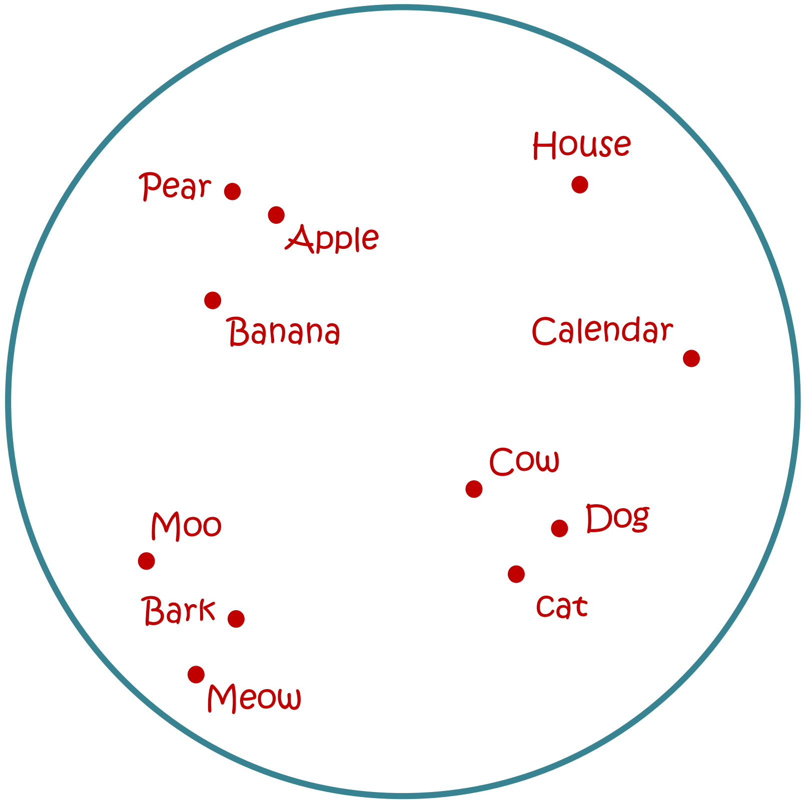
$$\log \sigma(w \cdot c) + \sum_{i=1}^{k} \mathbb{E}_{c_N \sim P_D}[\log \sigma(-w \cdot c_N)]$$

I will go to New York by plane .

$\underline{go}$ = w

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

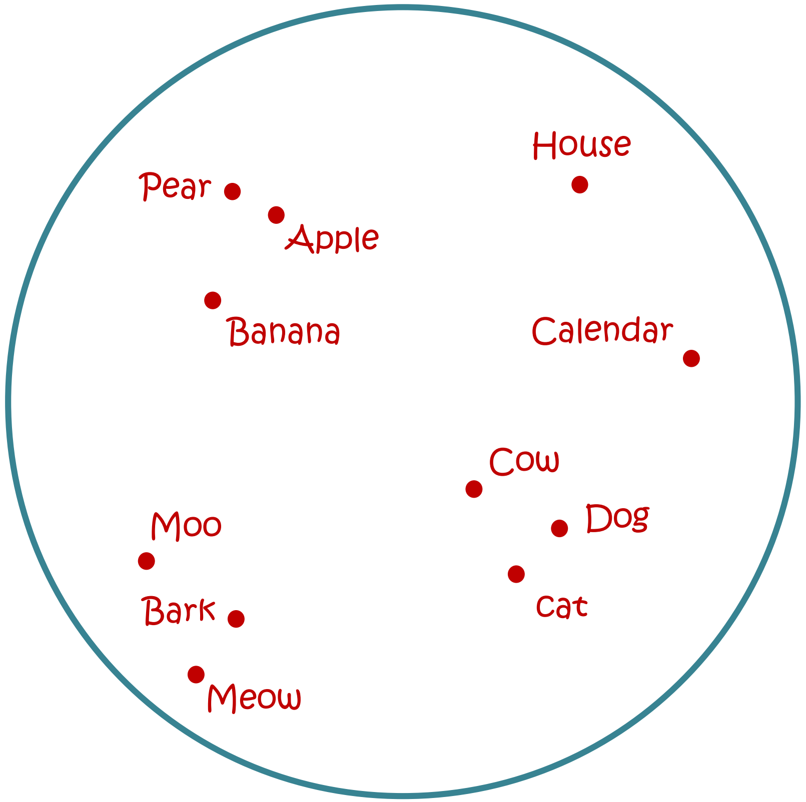
Skip-gram with negative sampling:

$$\log \sigma(w \cdot c) + \sum_{i=1}^{k} \mathbb{E}_{c_N \sim P_D}[\log \sigma(-w \cdot c_N)]$$

I will $\underset{w}{\underline{\text{go}}}$ to New York by plane .

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

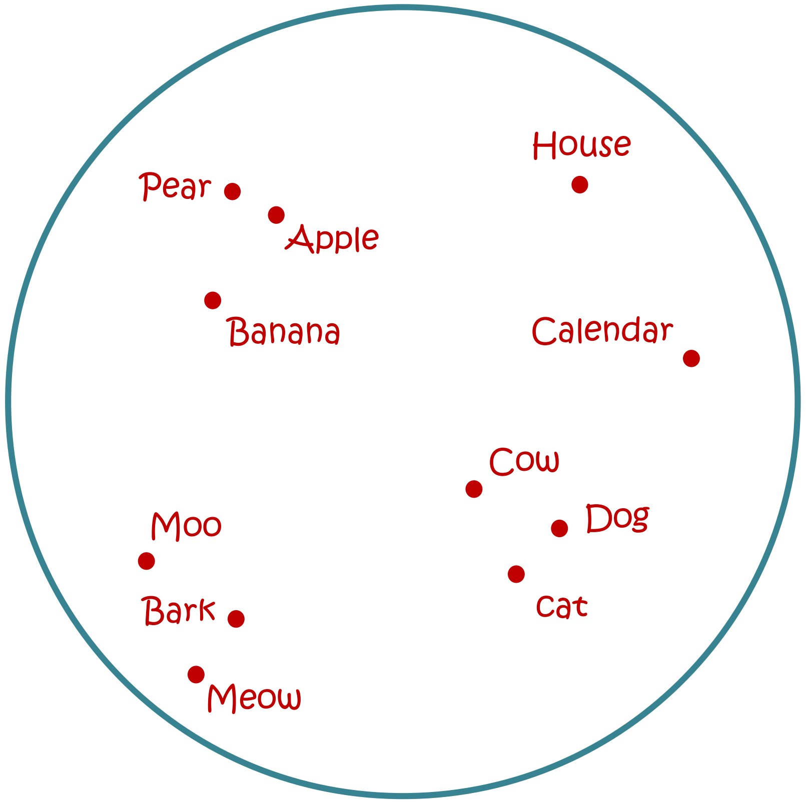
Skip-gram with negative sampling:

$$\log \sigma(w \cdot c) + \sum_{i=1}^{k} \mathbb{E}_{c_N \sim P_D}[\log \sigma(-w \cdot c_N)]$$

I will $\underbrace{\textit{go}}_{w}$ $\underbrace{\textit{to}}_{c}$ *New York by plane .*

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

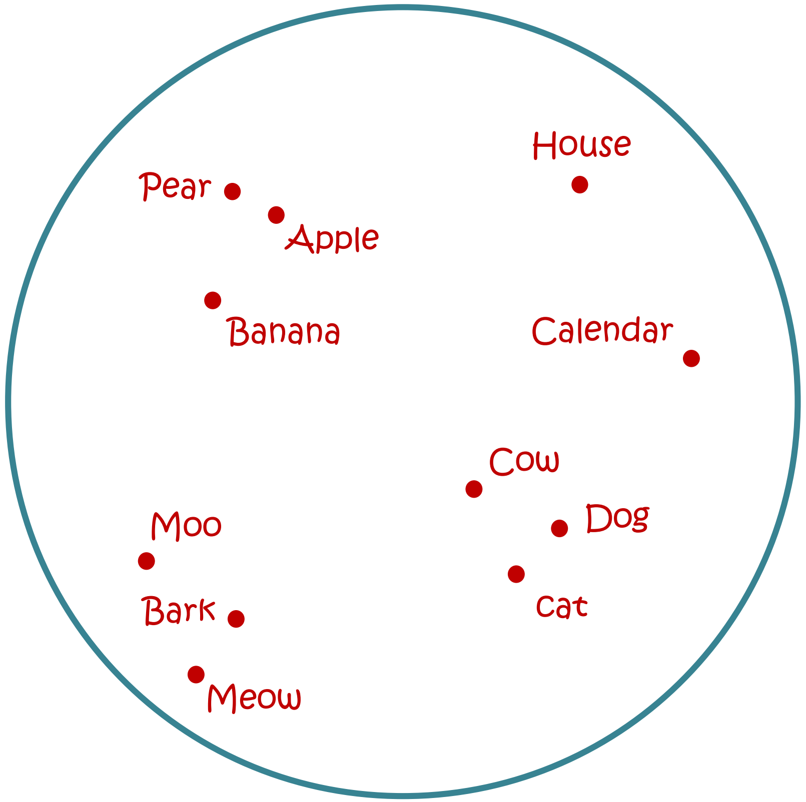
Skip-gram with negative sampling:

$$\log \sigma(w \cdot c) + \sum_{i=1}^{k} \mathbb{E}_{c_N \sim P_D}[\log \sigma(-w \cdot c_N)]$$

I will $\underset{w}{\underline{\textit{go}}}$ *to* $\underset{c}{\underline{\textit{New}}}$ *York by plane .*

Semantic space

- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Embeddings

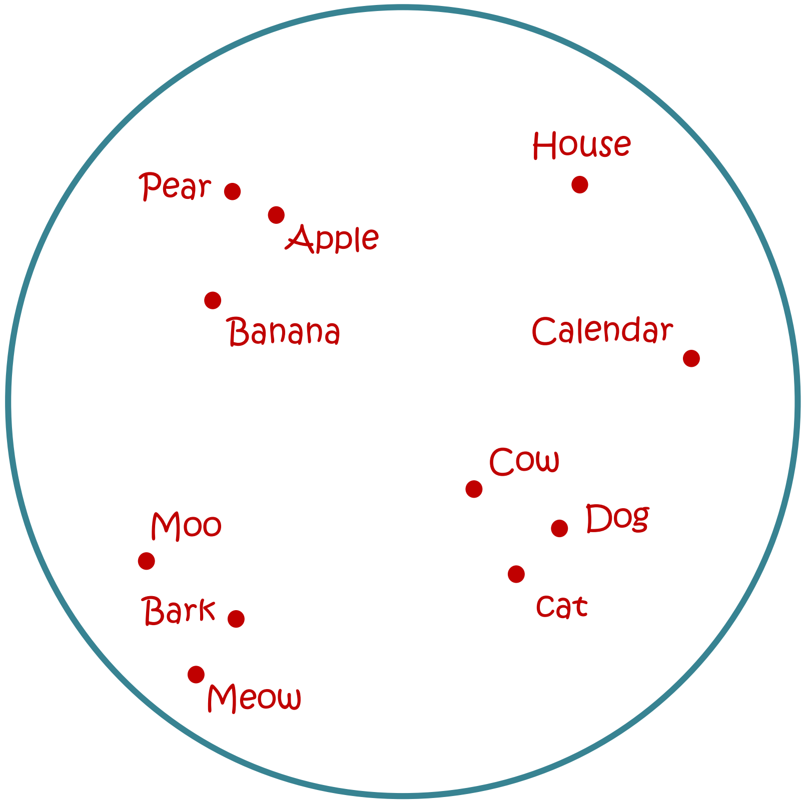
Skip-gram with negative sampling:

$$\log \sigma(w \cdot c) + \sum_{i=1}^{k} \mathbb{E}_{c_N \sim P_D}[\log \sigma(-w \cdot c_N)]$$

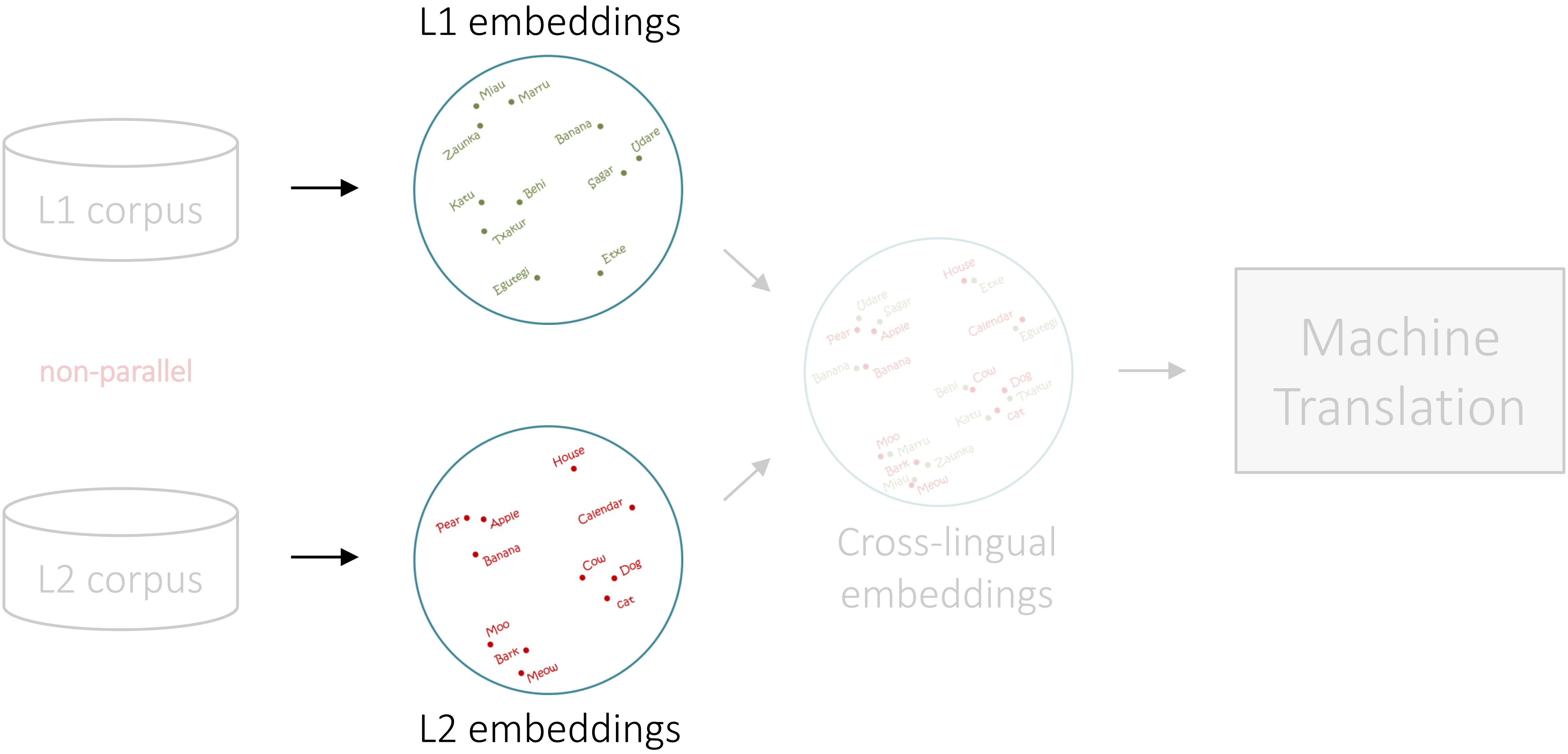
$$I \;\; will \;\; \underbrace{go}_{w} \;\; to \;\; New \;\; \underbrace{York}_{c} \;\; by \;\; plane \;\; .$$

Semantic space

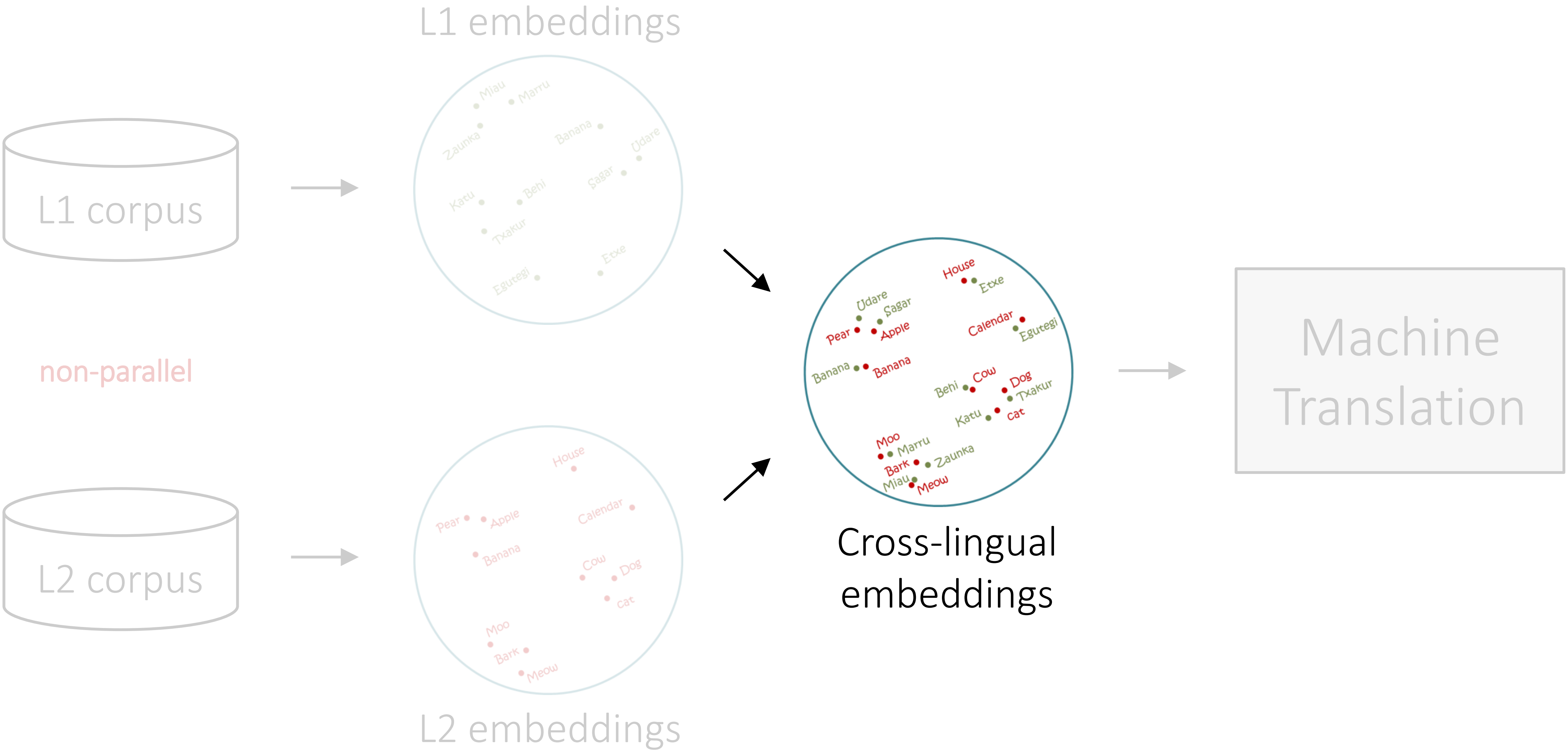
- Words
- Meaningful distances
- Meaningful relations
- 300 dimensions
- Machine learning / linear algebra from co-occurrence counts



Outline

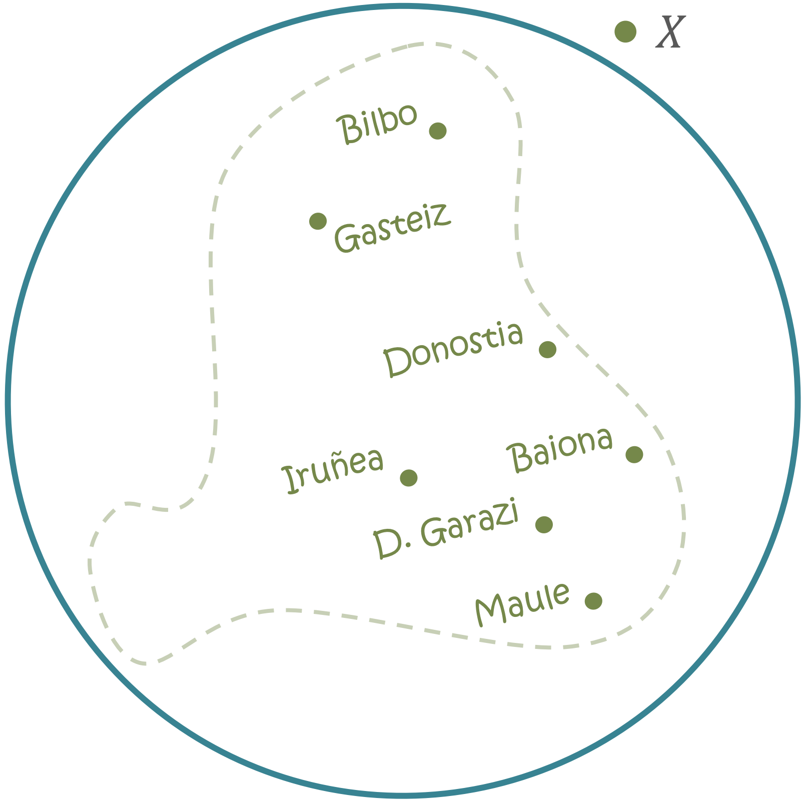

L1 embeddings
L1 corpus
Miau
Marru
Zaunka
Banana
Udare
Katu
Behi
Sagar
Txakur
Etxe
Egutegi
non-parallel
L2 corpus
House
Pear
Apple
Calendar
Banana
Cow
Dog
cat
Moo
Bark
Meow
L2 embeddings
Cross-lingual embeddings
Machine Translation

Outline

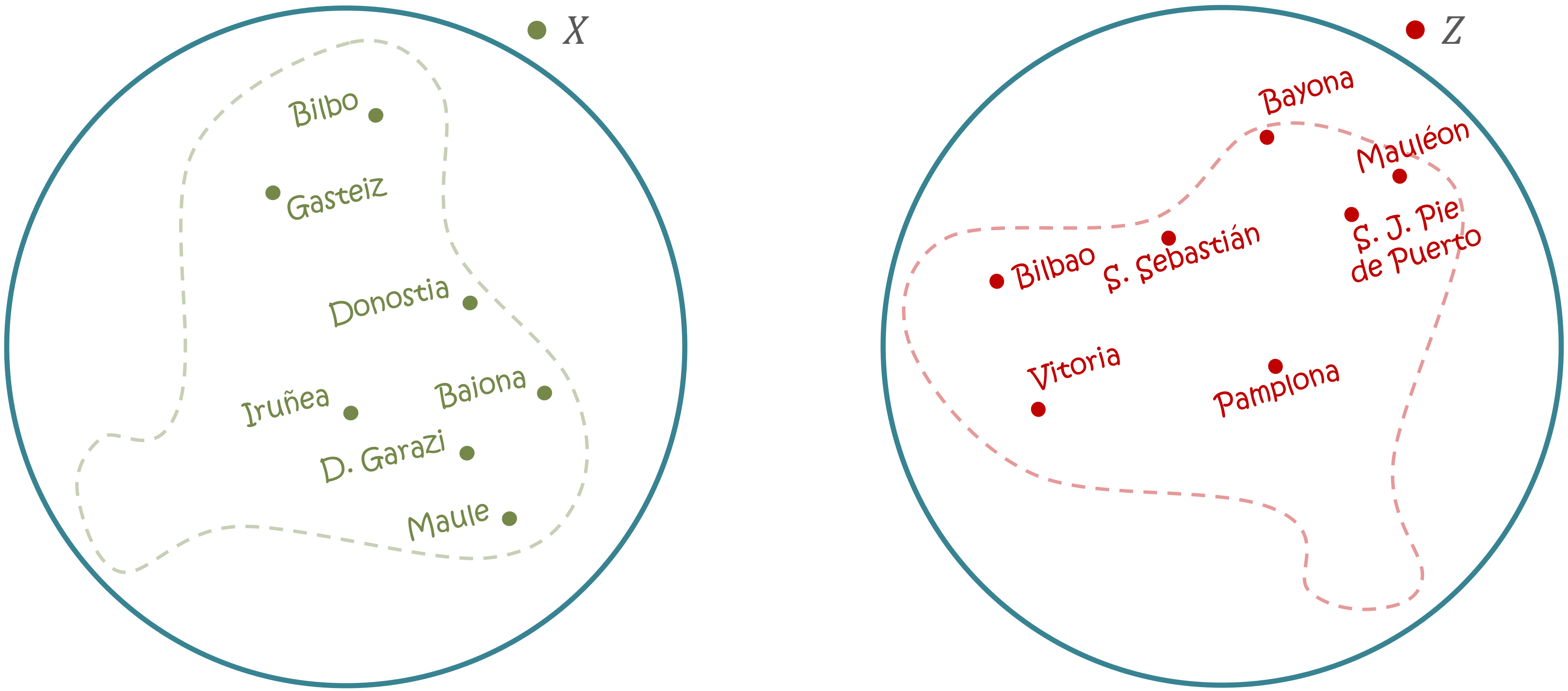

L1 embeddings
L1 corpus
non-parallel
L2 corpus
L2 embeddings
Miau
Marru
Zaunka
Banana
Udare
Sagar
Katu
Behi
Txakur
Egutegi
Etxe
House
Pear
Apple
Banana
Calendar
Cow
Dog
Cat
Moo
Bark
Meow
Cross-lingual embeddings
Machine Translation

Cross-lingual embedding mappings

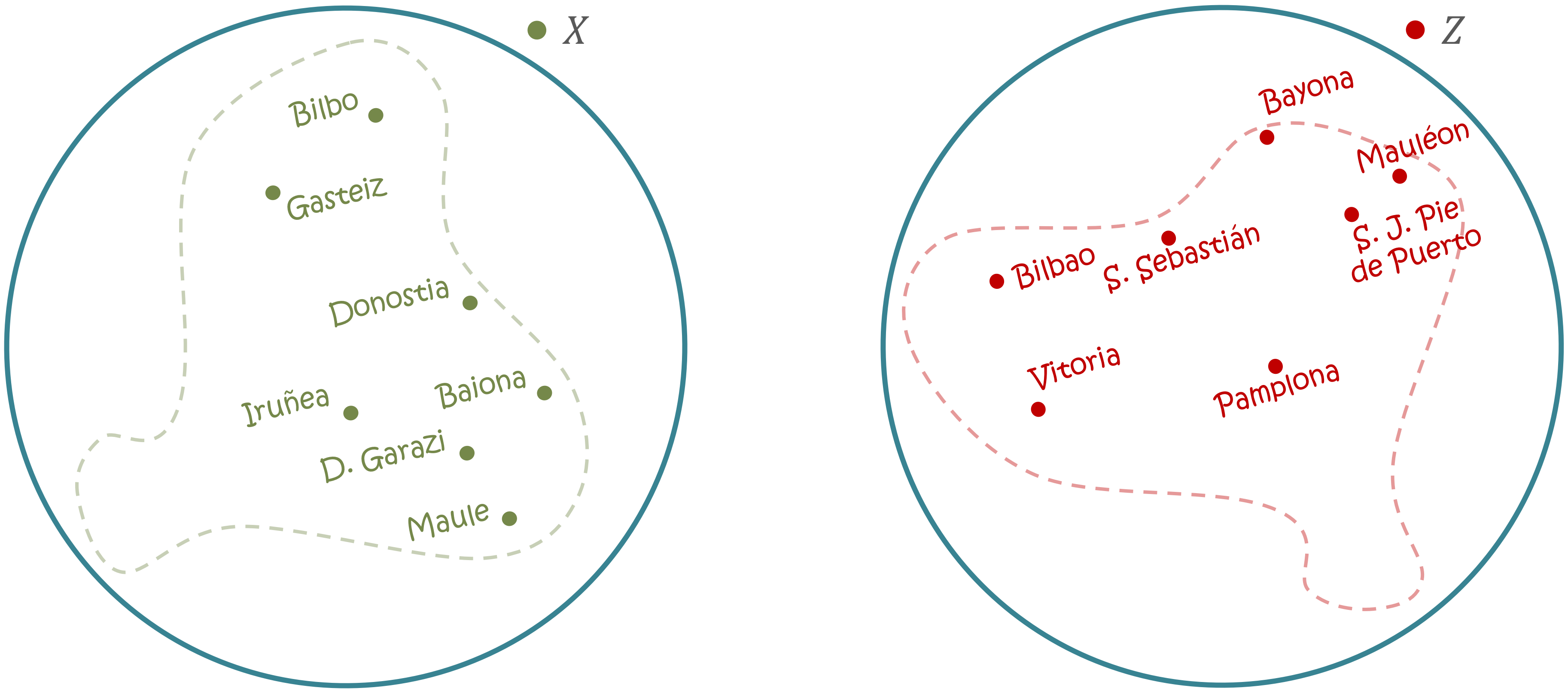
Cross-lingual embedding mappings



Cross-lingual embedding mappings

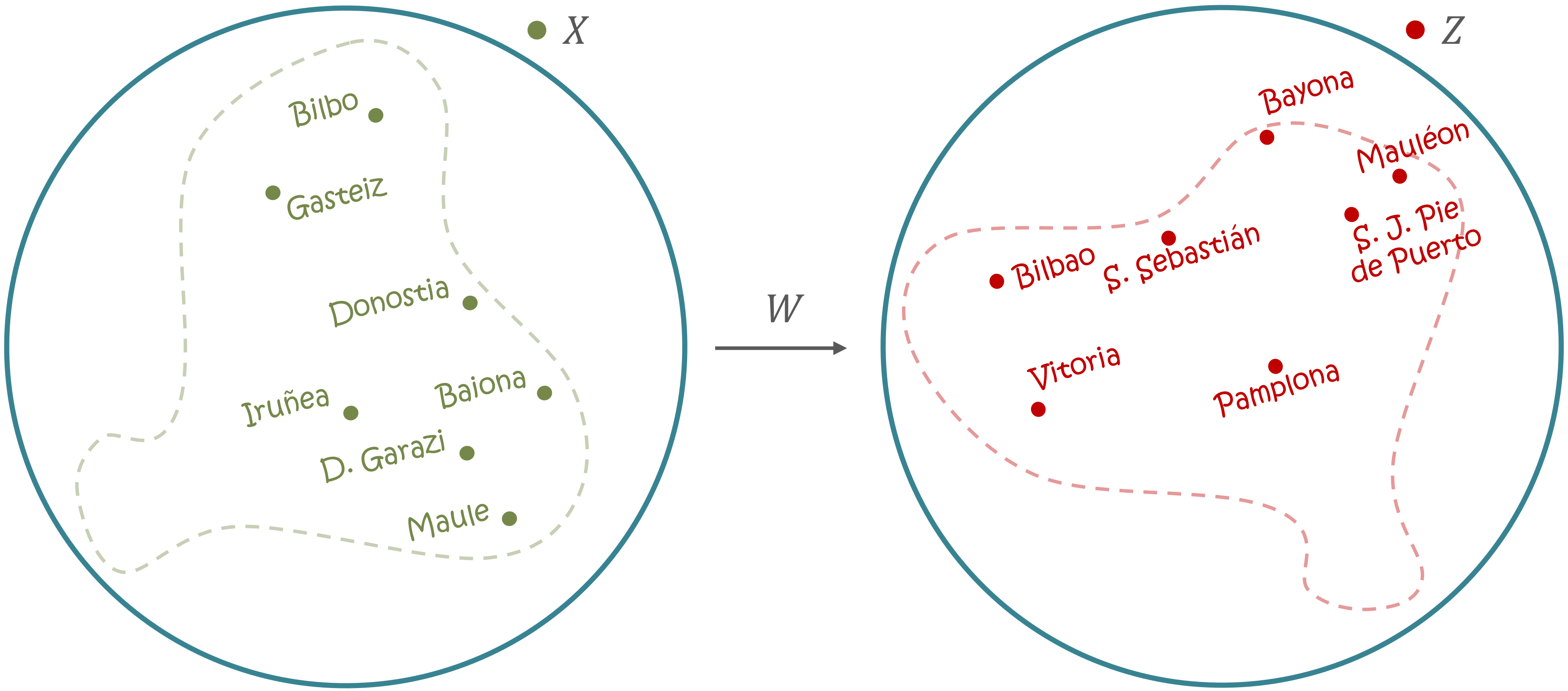


Cross-lingual embedding mappings



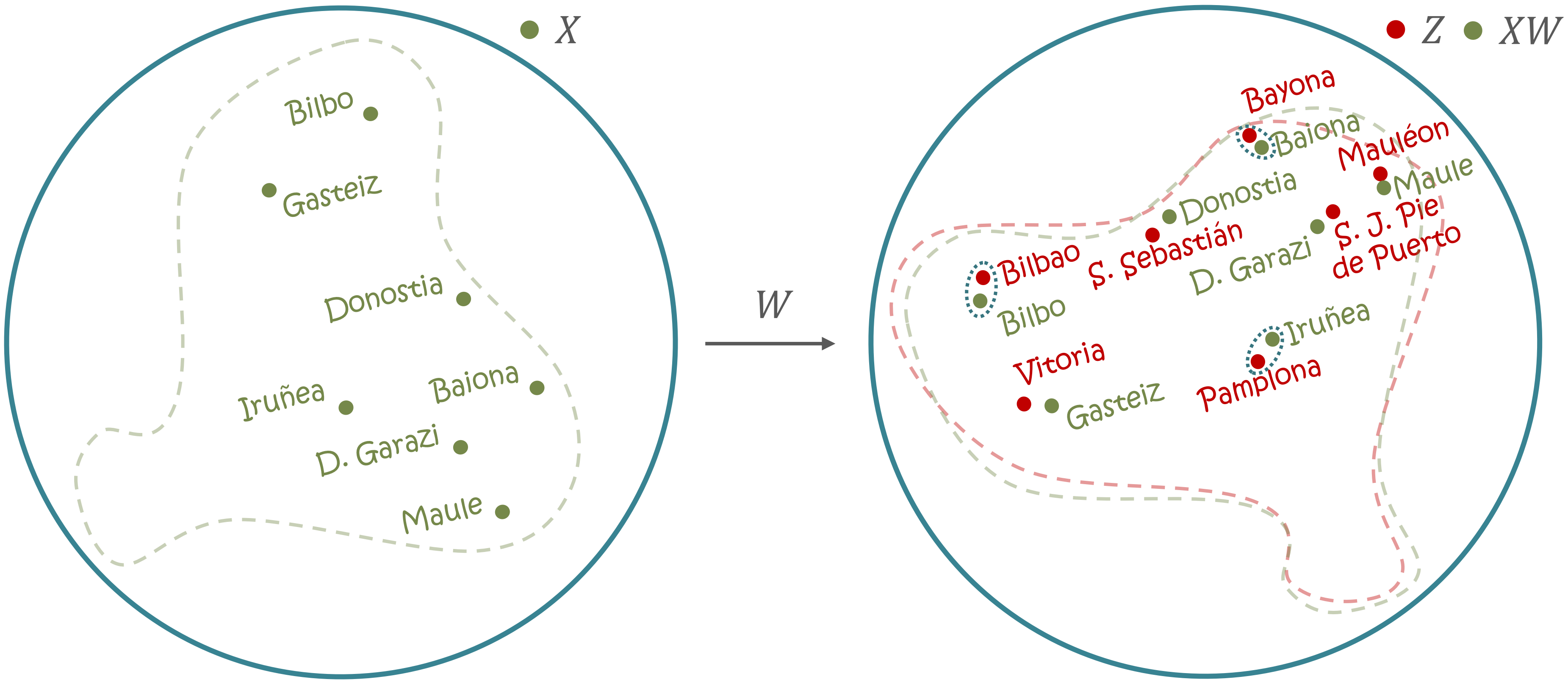
Bilbo	Bilbao
Baiona	Bayona
Iruñea	Pamplona

Cross-lingual embedding mappings



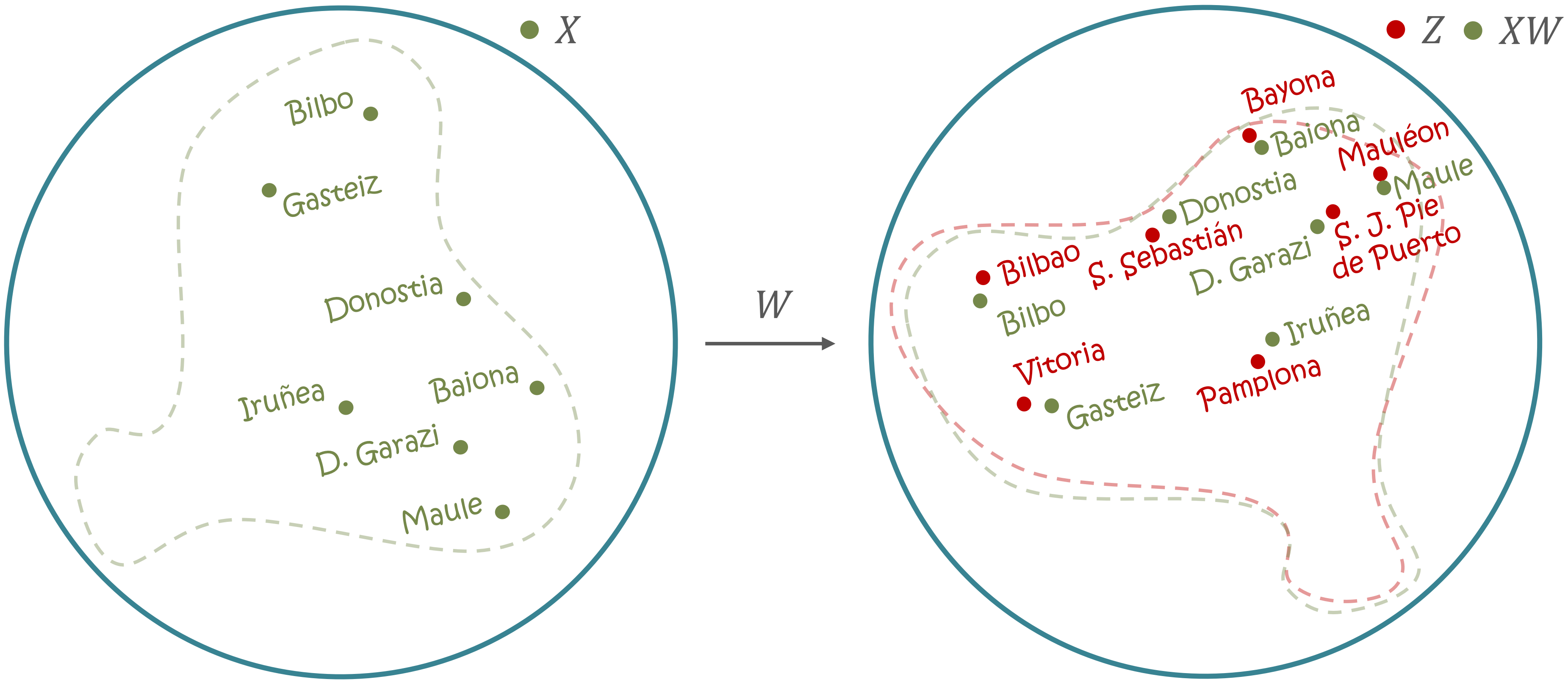
Bilbo	Bilbao
Baiona	Bayona
Iruñea	Pamplona

Cross-lingual embedding mappings



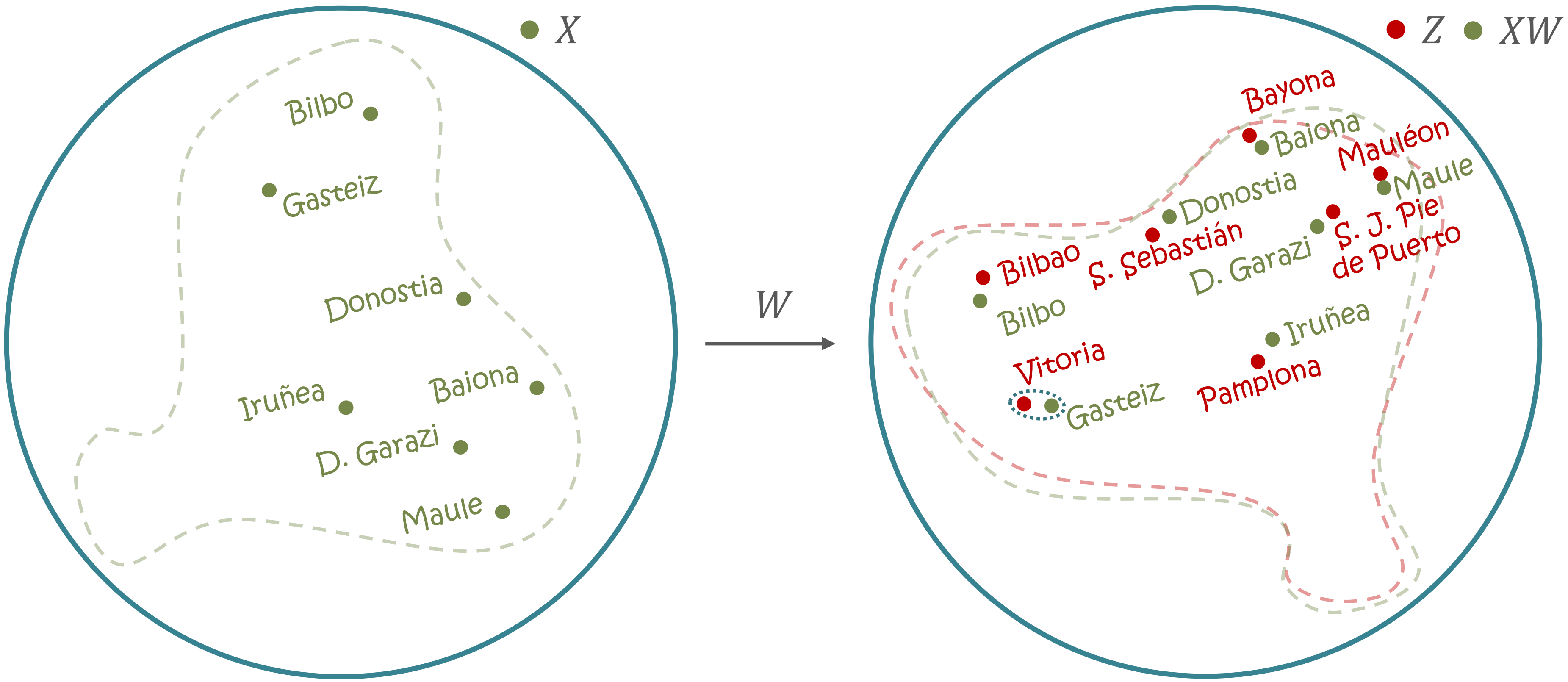
Bilbo	Bilbao
Baiona	Bayona
Iruñea	Pamplona

Cross-lingual embedding mappings



Bilbo	Bilbao
Baiona	Bayona
Iruñea	Pamplona

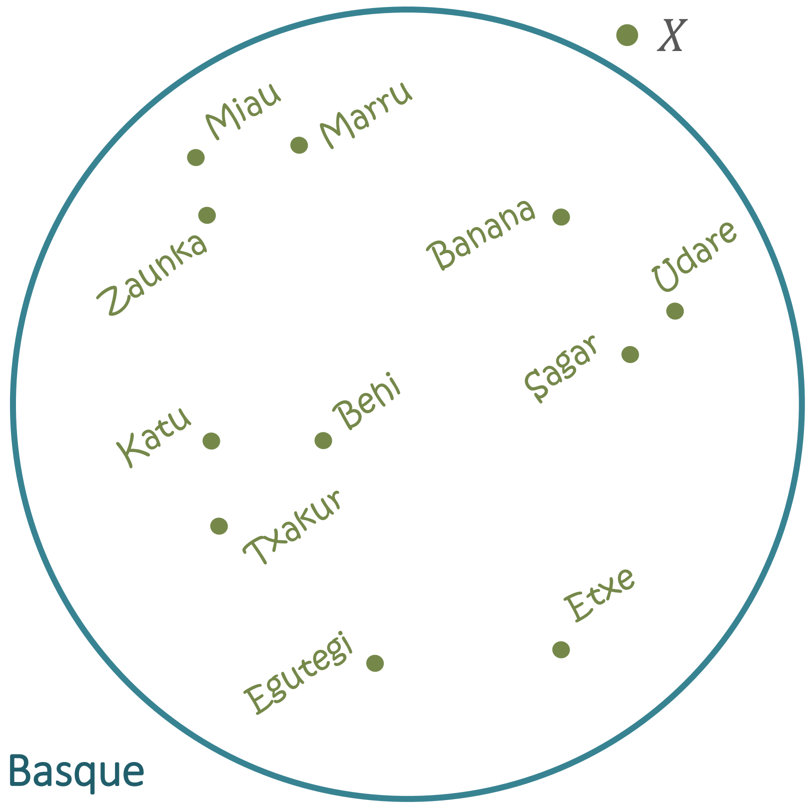
Cross-lingual embedding mappings



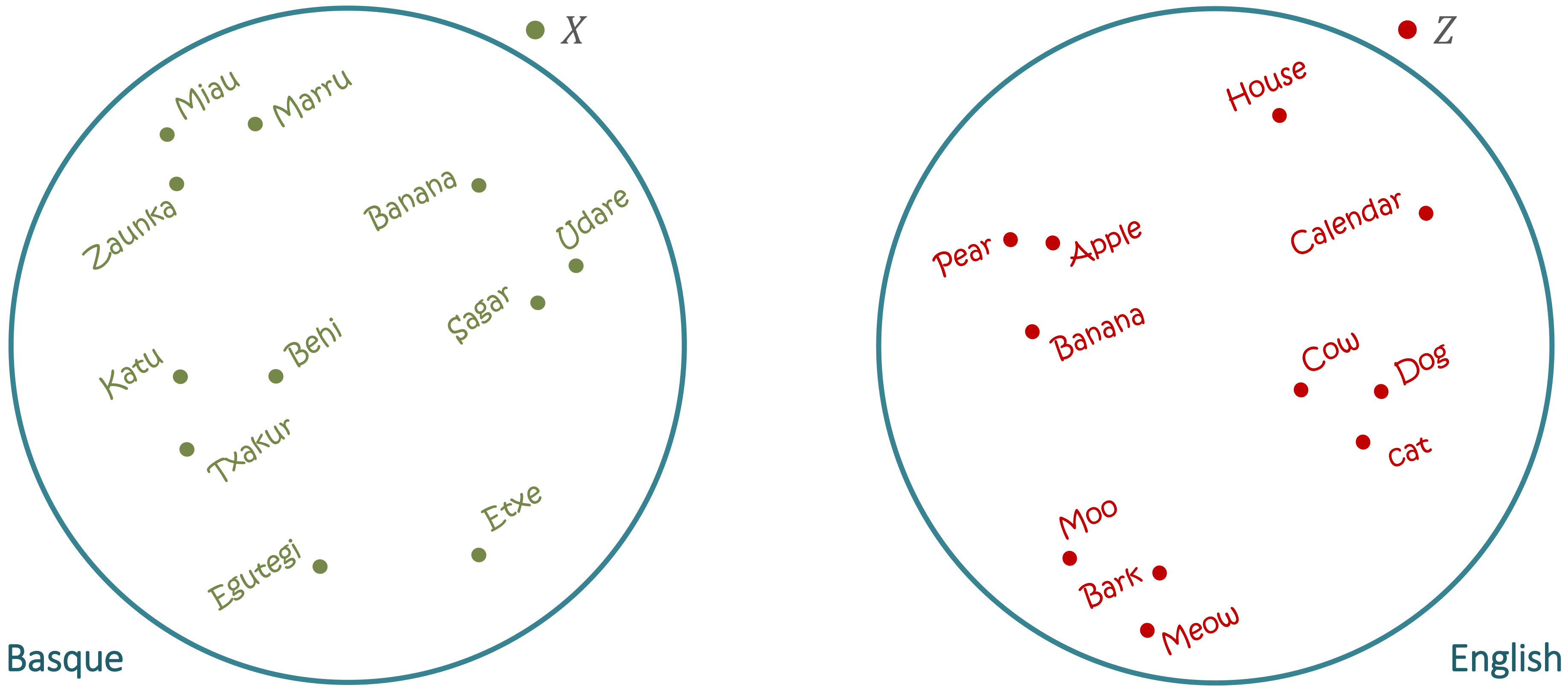
Bilbo	Bilbao
Baiona	Bayona
Iruñea	Pamplona

Cross-lingual embedding mappings

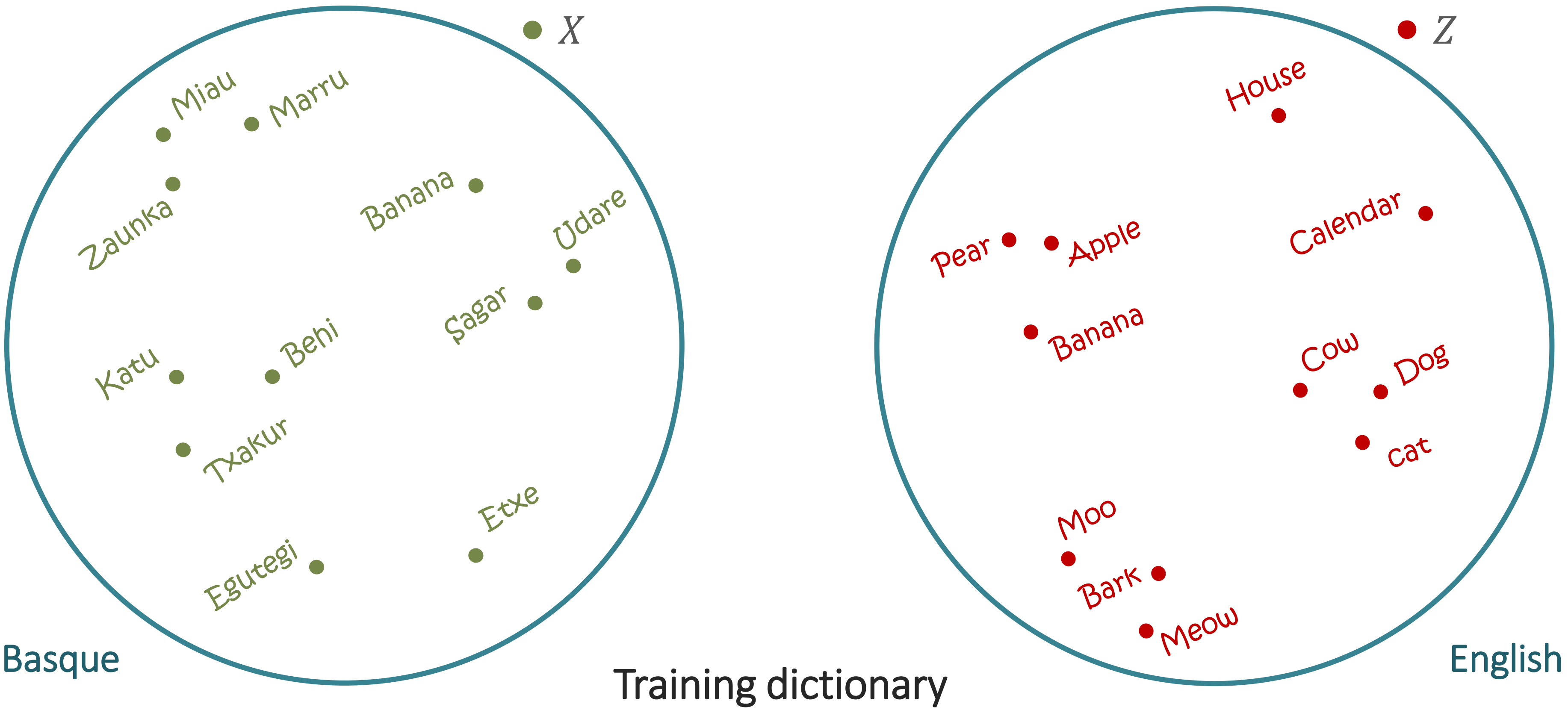
Cross-lingual embedding mappings



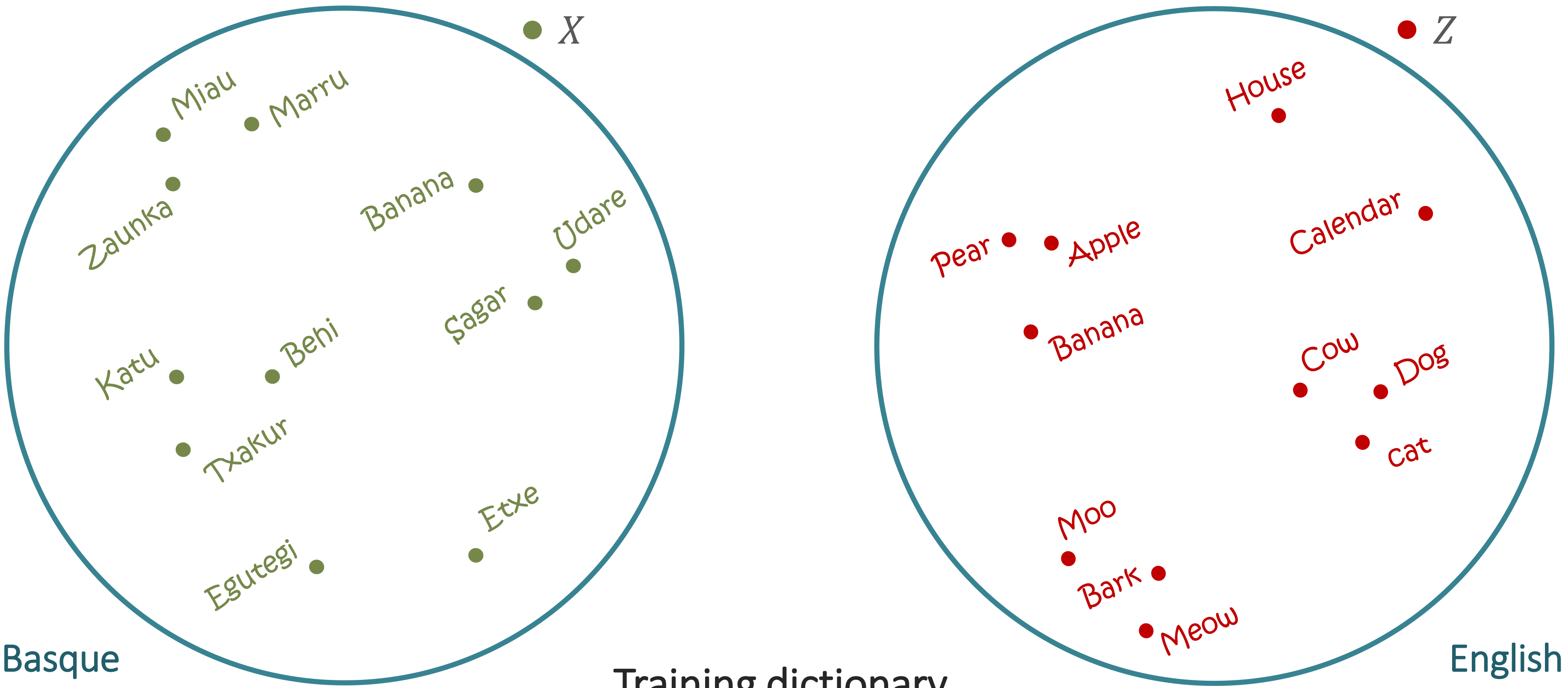
Cross-lingual embedding mappings



Cross-lingual embedding mappings



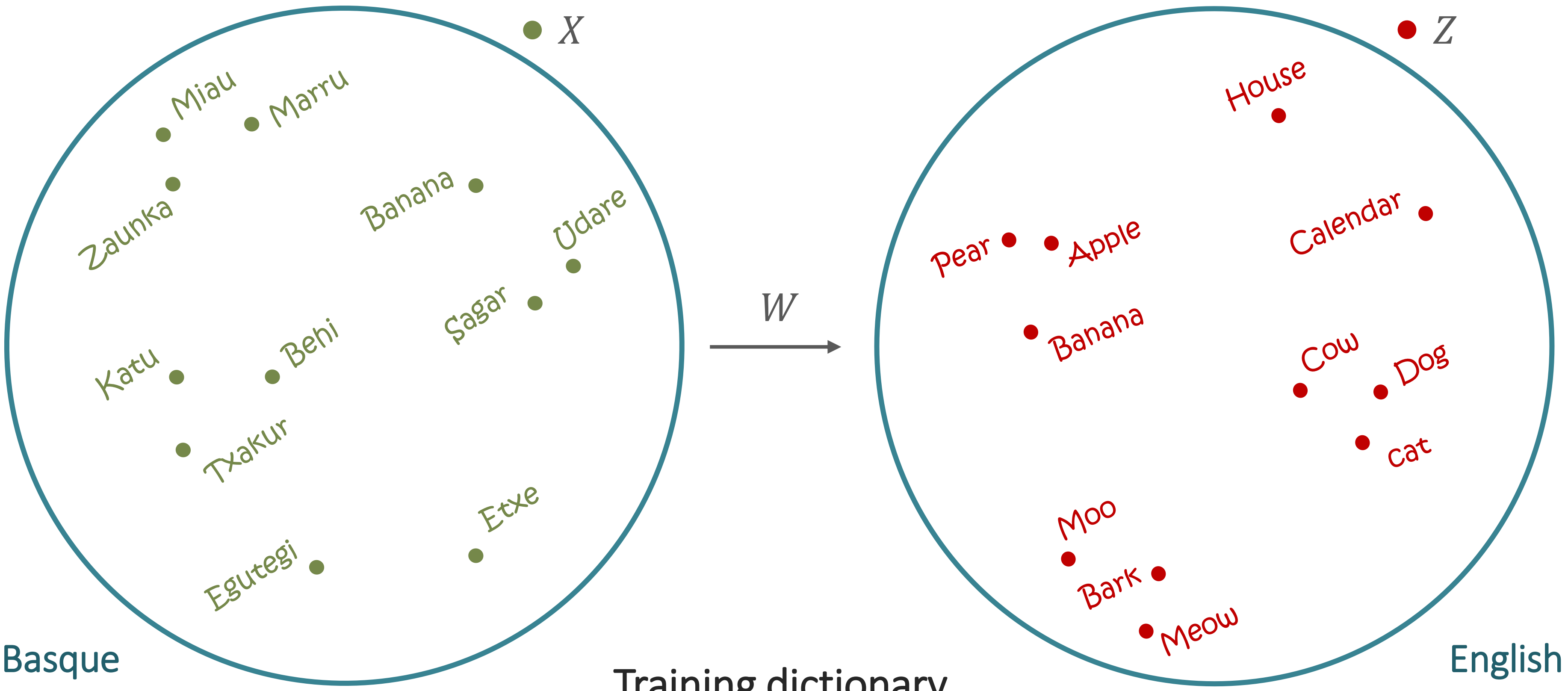
Cross-lingual embedding mappings



Training dictionary

Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

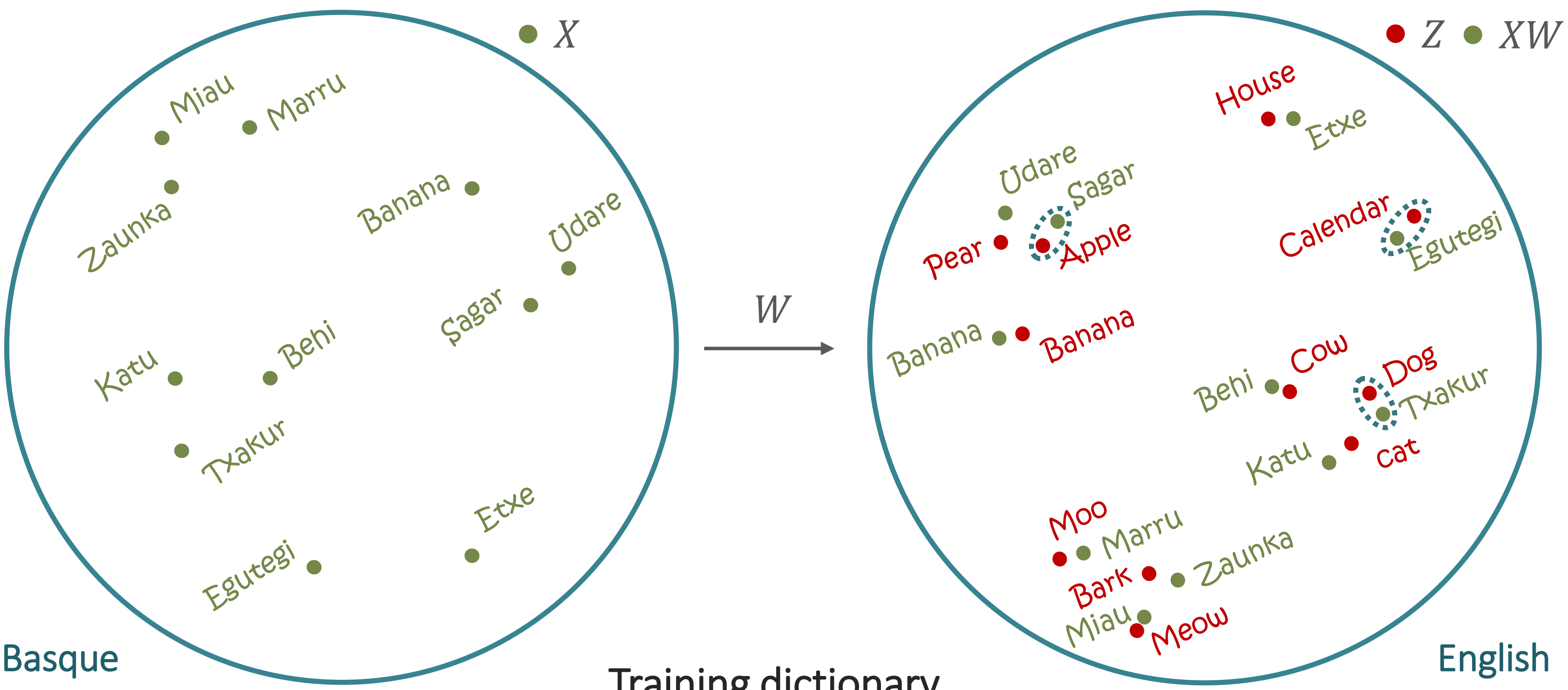
Cross-lingual embedding mappings



Training dictionary

Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

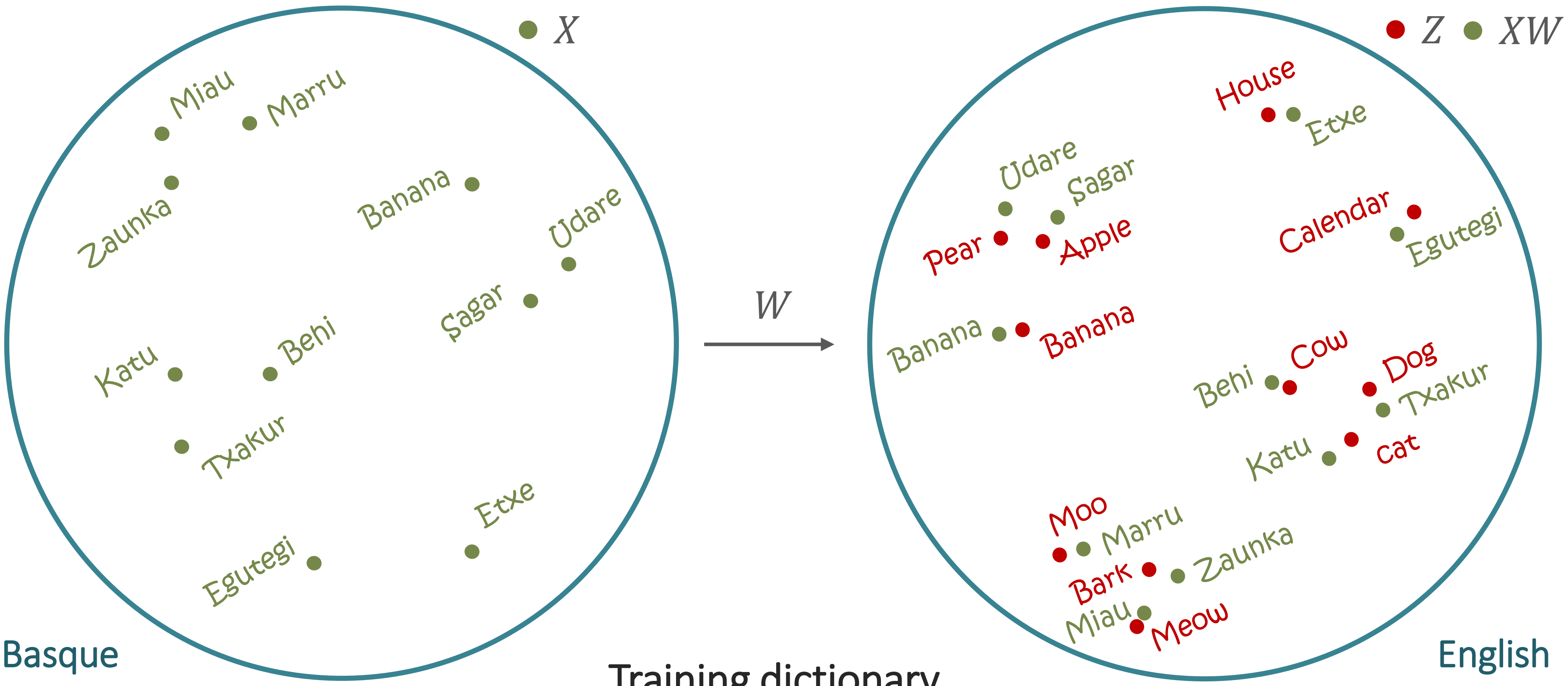
Cross-lingual embedding mappings



Training dictionary

Basque	English
Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

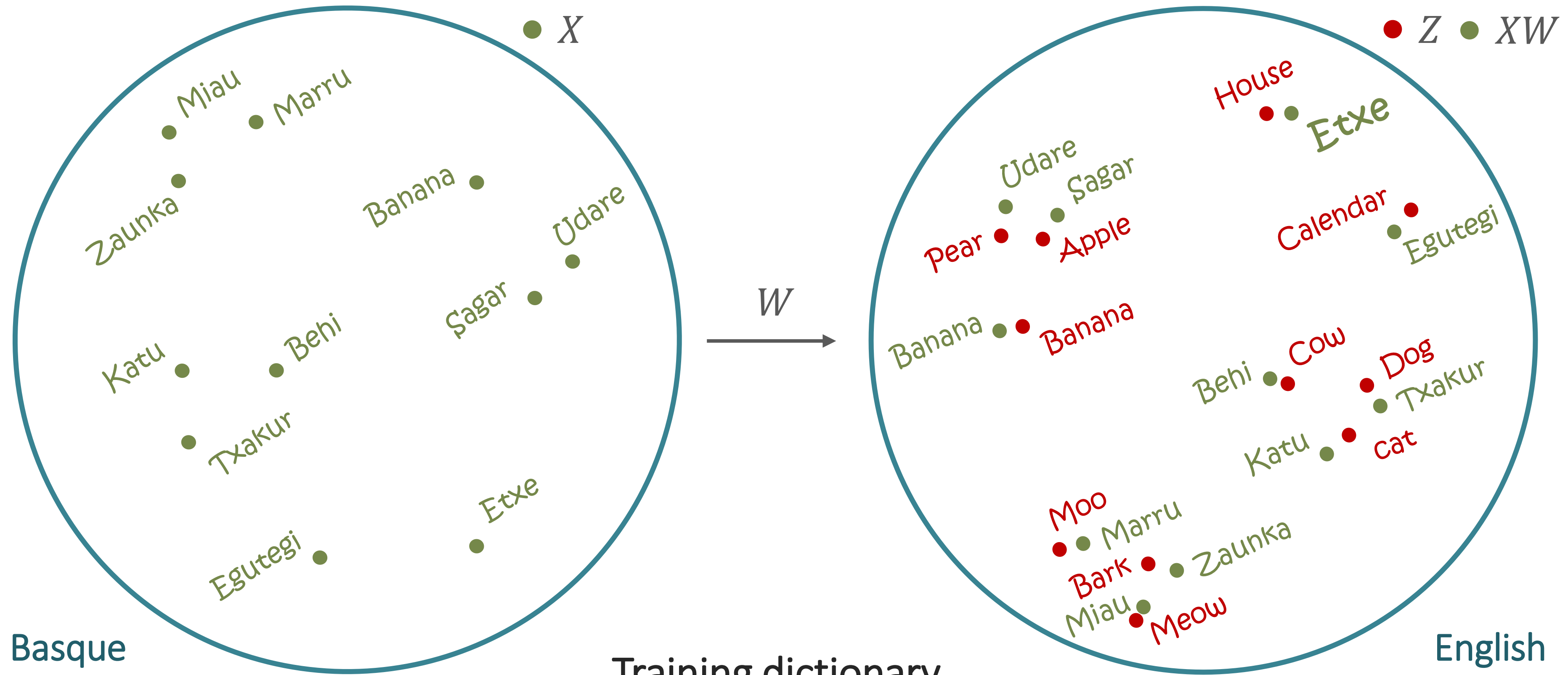
Cross-lingual embedding mappings



Training dictionary

Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

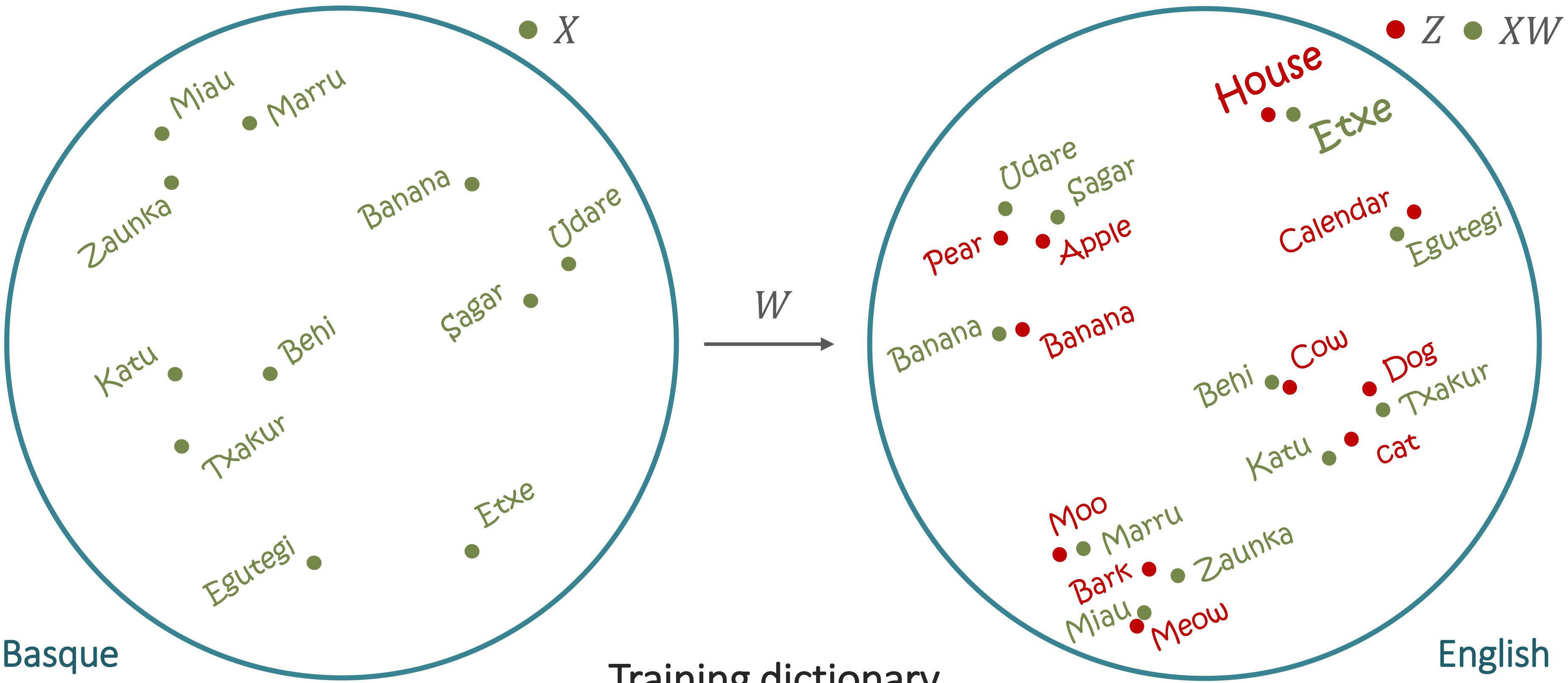
Cross-lingual embedding mappings



Training dictionary

Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

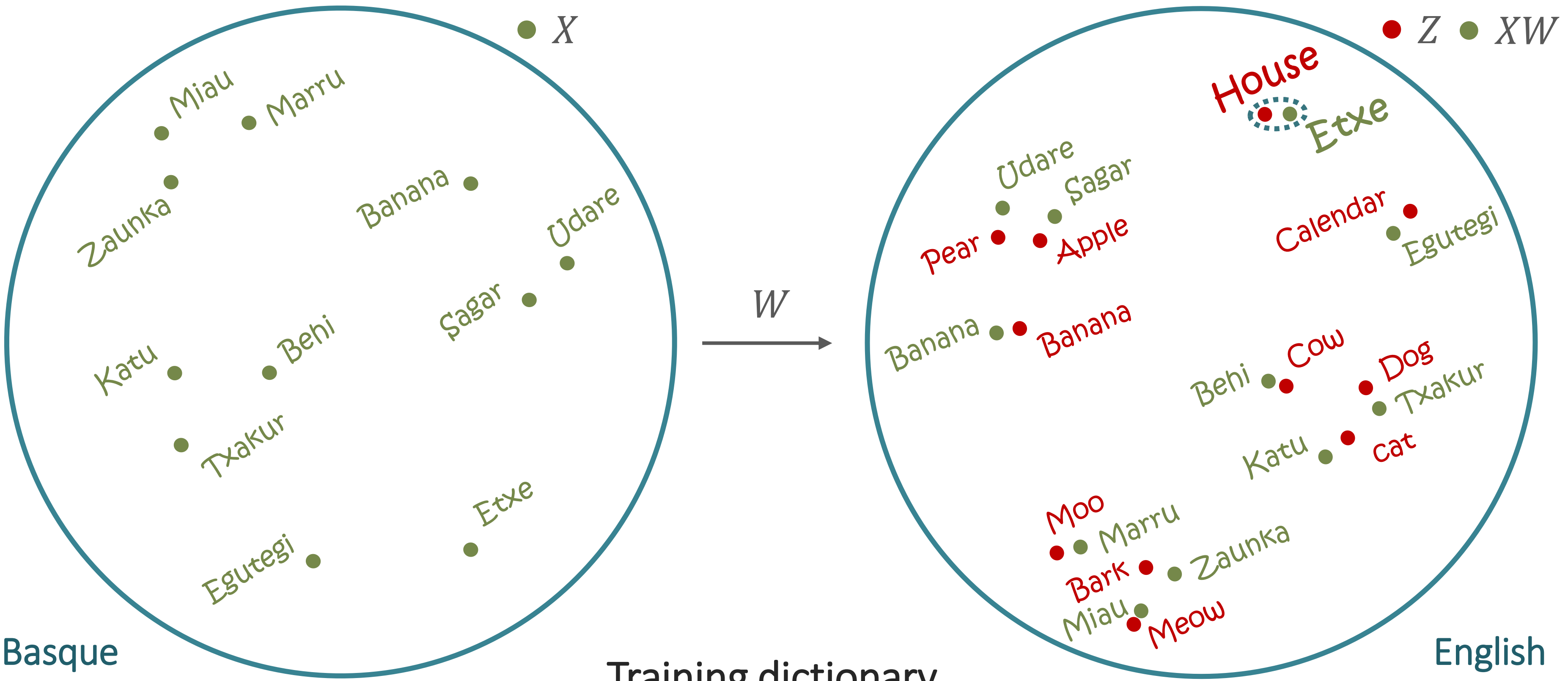
Cross-lingual embedding mappings



Training dictionary

Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

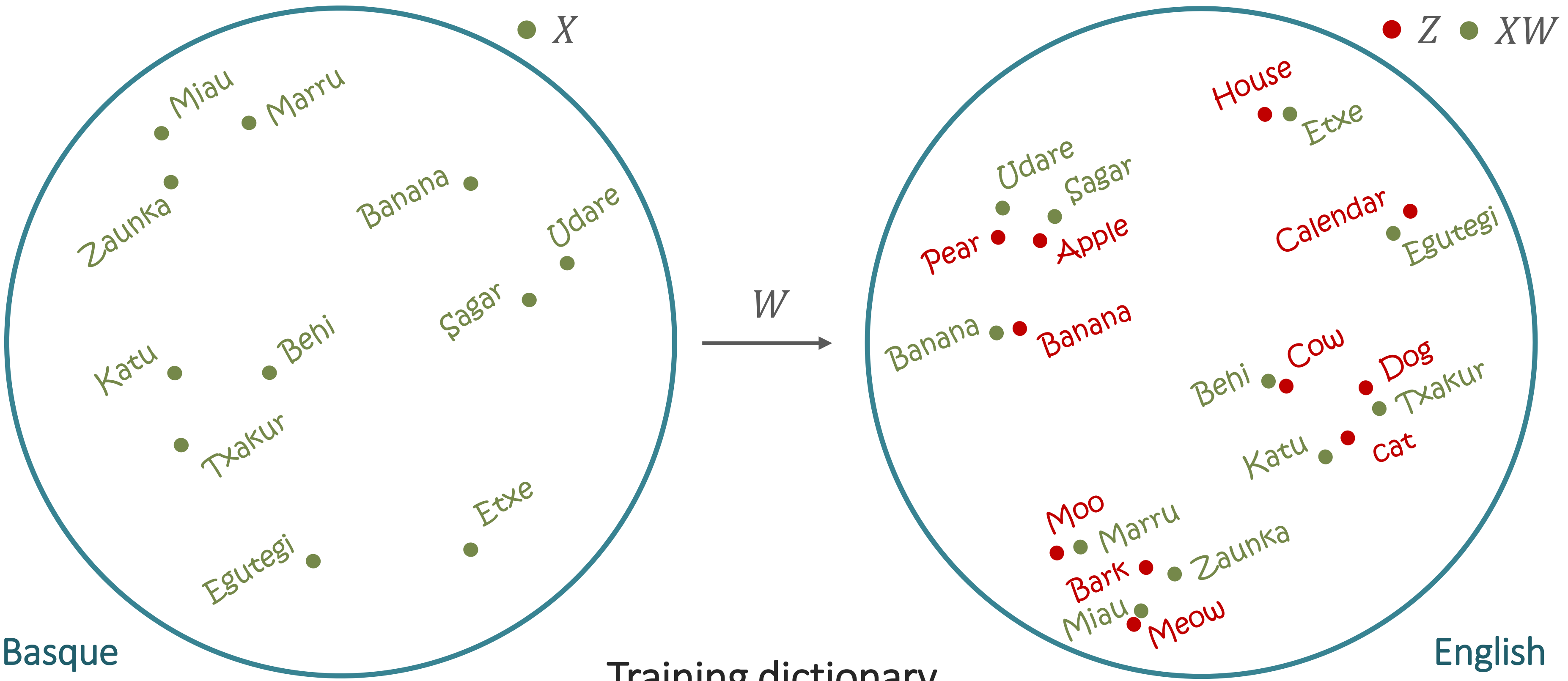
Cross-lingual embedding mappings



Training dictionary

Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

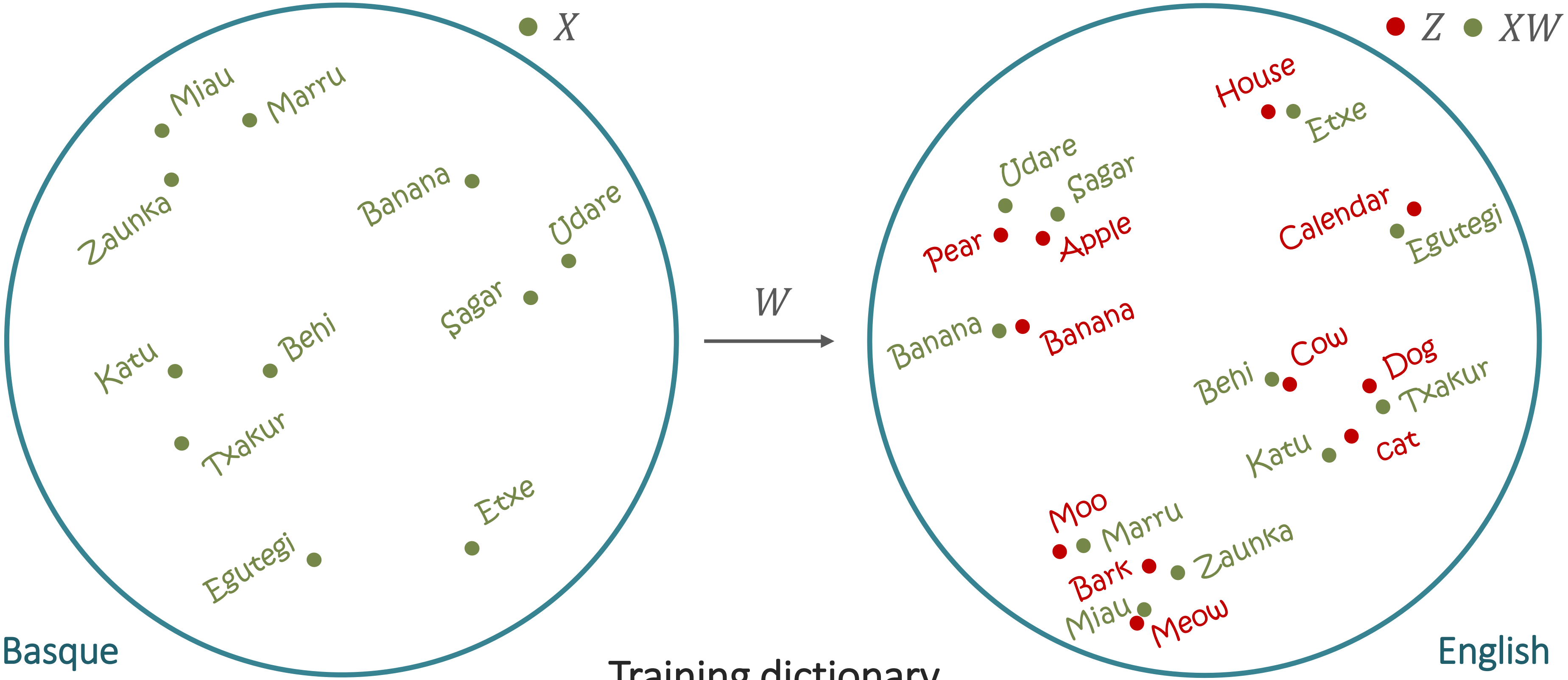
Cross-lingual embedding mappings



Training dictionary

Txakur	Dog
Sagar	Apple
⋮	⋮
Egutegi	Calendar

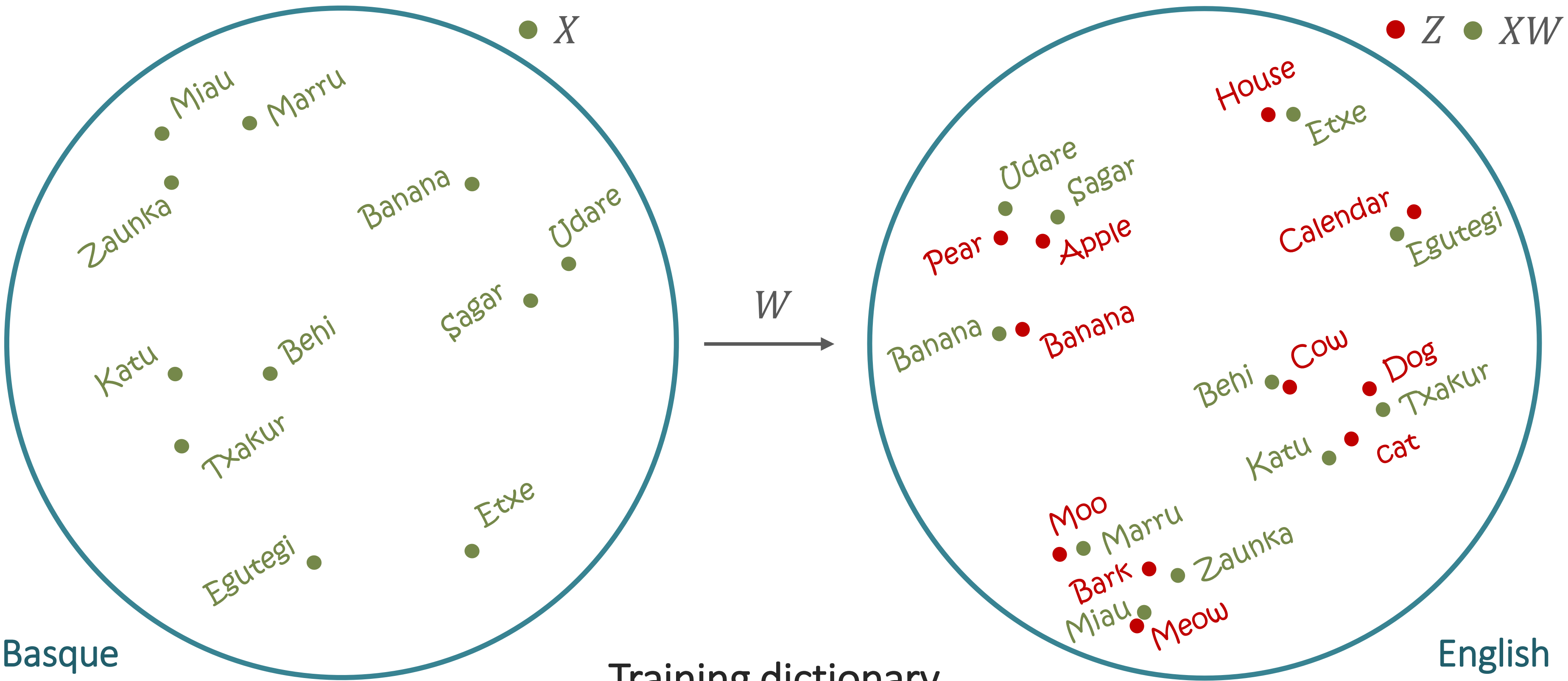
Cross-lingual embedding mappings



Training dictionary

$$\begin{array}{r} \text{Txakur} \\ \text{Sagar} \\ \vdots \\ \text{Egutegi} \end{array} \begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_{n,*} \end{bmatrix} \quad \Bigg| \quad \begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_{n,*} \end{bmatrix} \begin{array}{l} \text{Dog} \\ \text{Apple} \\ \vdots \\ \text{Calendar} \end{array}$$

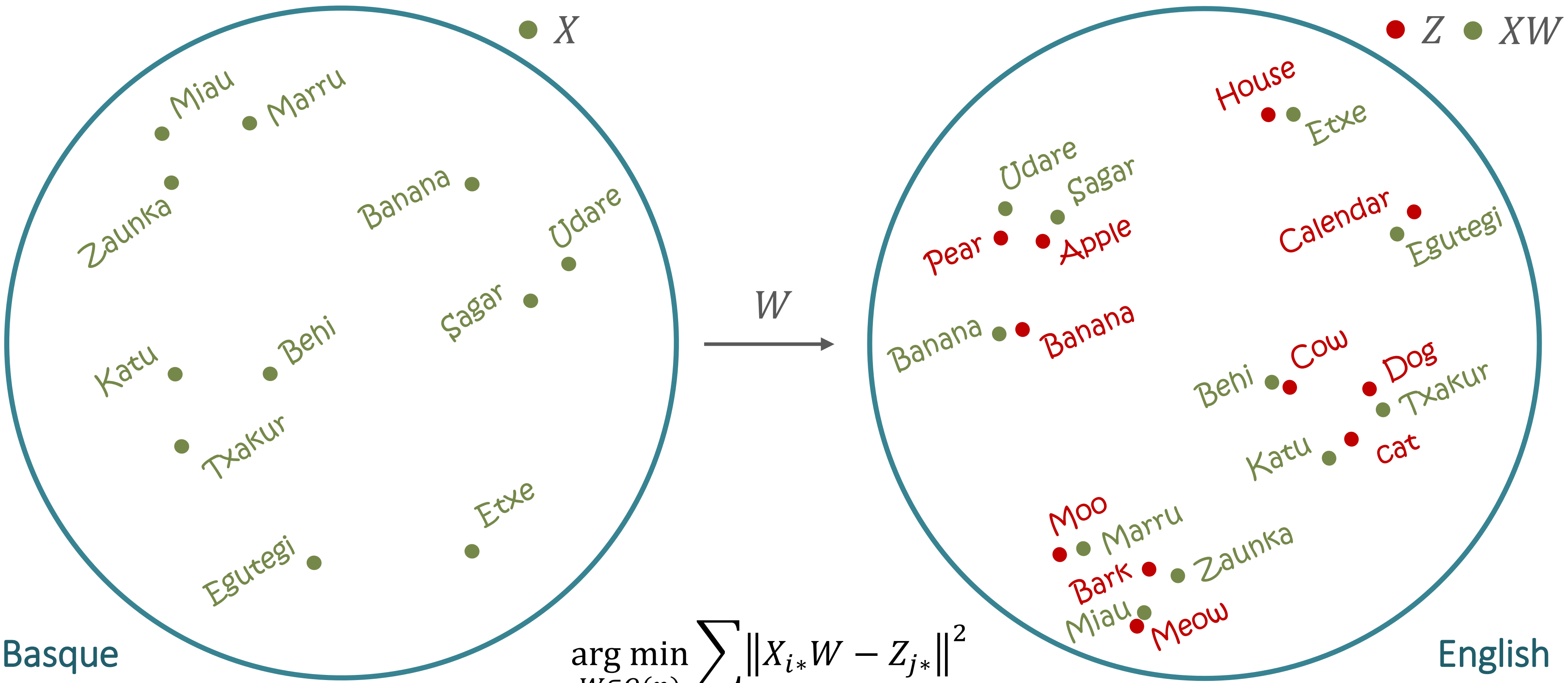
Cross-lingual embedding mappings



Training dictionary

$$
\begin{array}{r}
\text{Txakur} \\ \text{Sagar} \\ \vdots \\ \text{Egutegi}
\end{array}
\begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_{n,*} \end{bmatrix}
[W] \approx
\begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_{n,*} \end{bmatrix}
\begin{array}{l}
\text{Dog} \\ \text{Apple} \\ \vdots \\ \text{Calendar}
\end{array}
$$

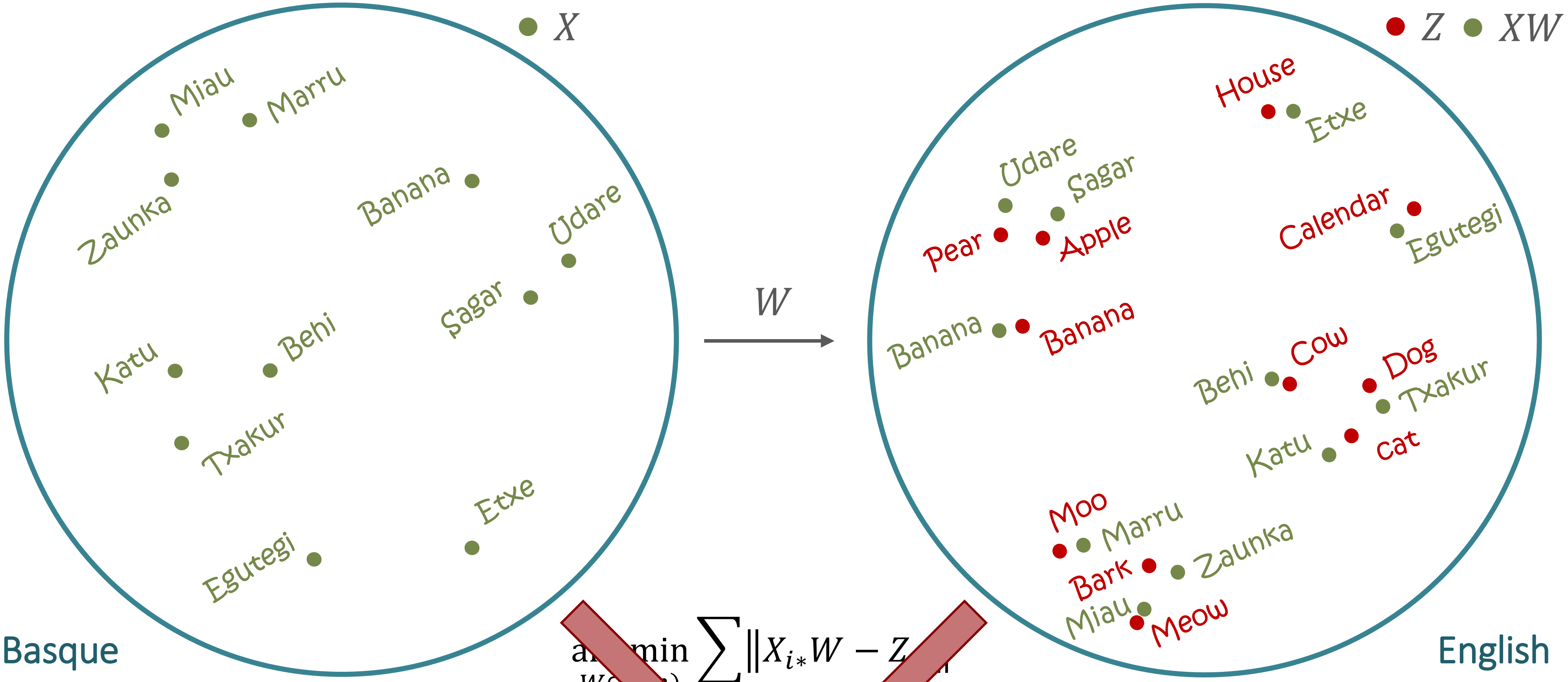
Cross-lingual embedding mappings



$$\underset{W \in O(n)}{\arg\min} \sum_i \left\| X_{i*} W - Z_{j*} \right\|^2$$

$$\begin{array}{r} \text{Txakur} \\ \text{Sagar} \\ \vdots \\ \text{Egutegi} \end{array} \begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_{n,*} \end{bmatrix} [W] \approx \begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_{n,*} \end{bmatrix} \begin{array}{l} \text{Dog} \\ \text{Apple} \\ \vdots \\ \text{Calendar} \end{array}$$

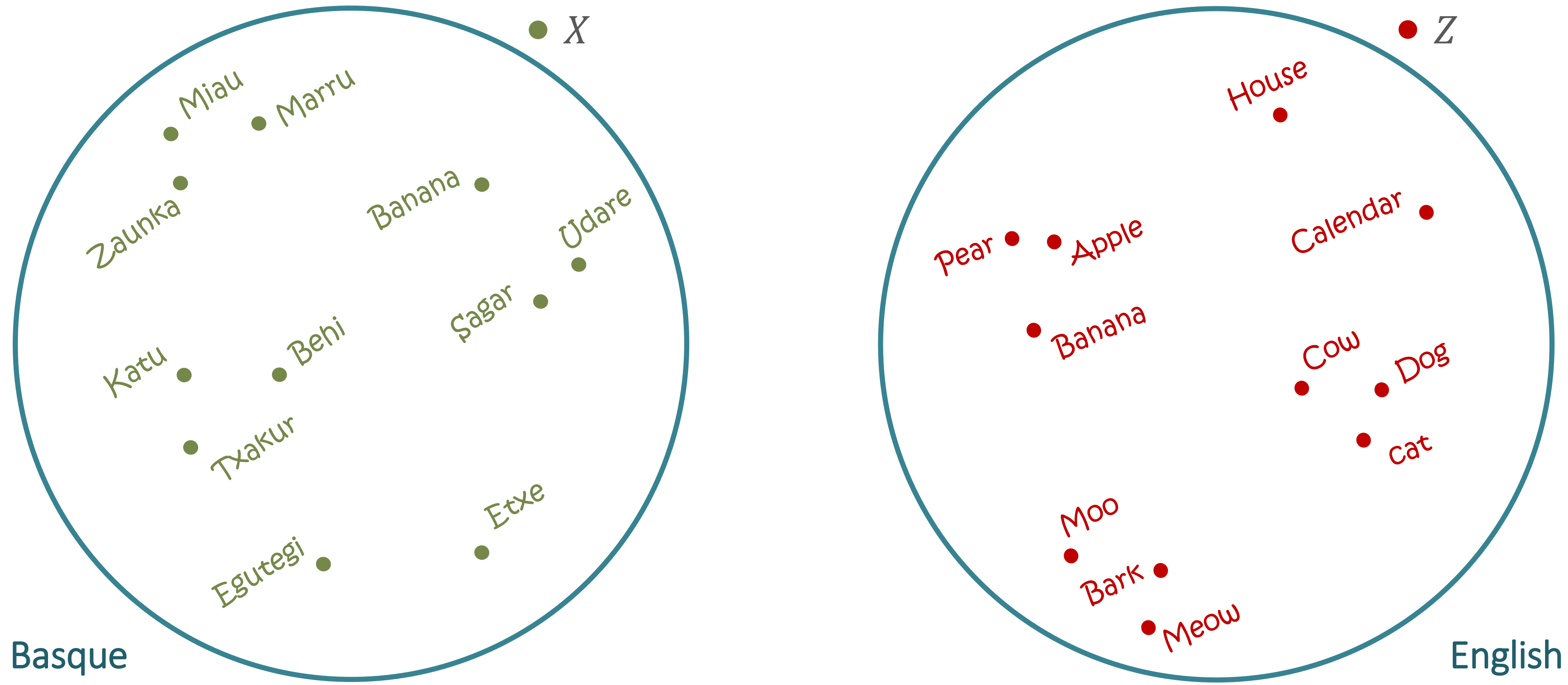
Cross-lingual embedding mappings



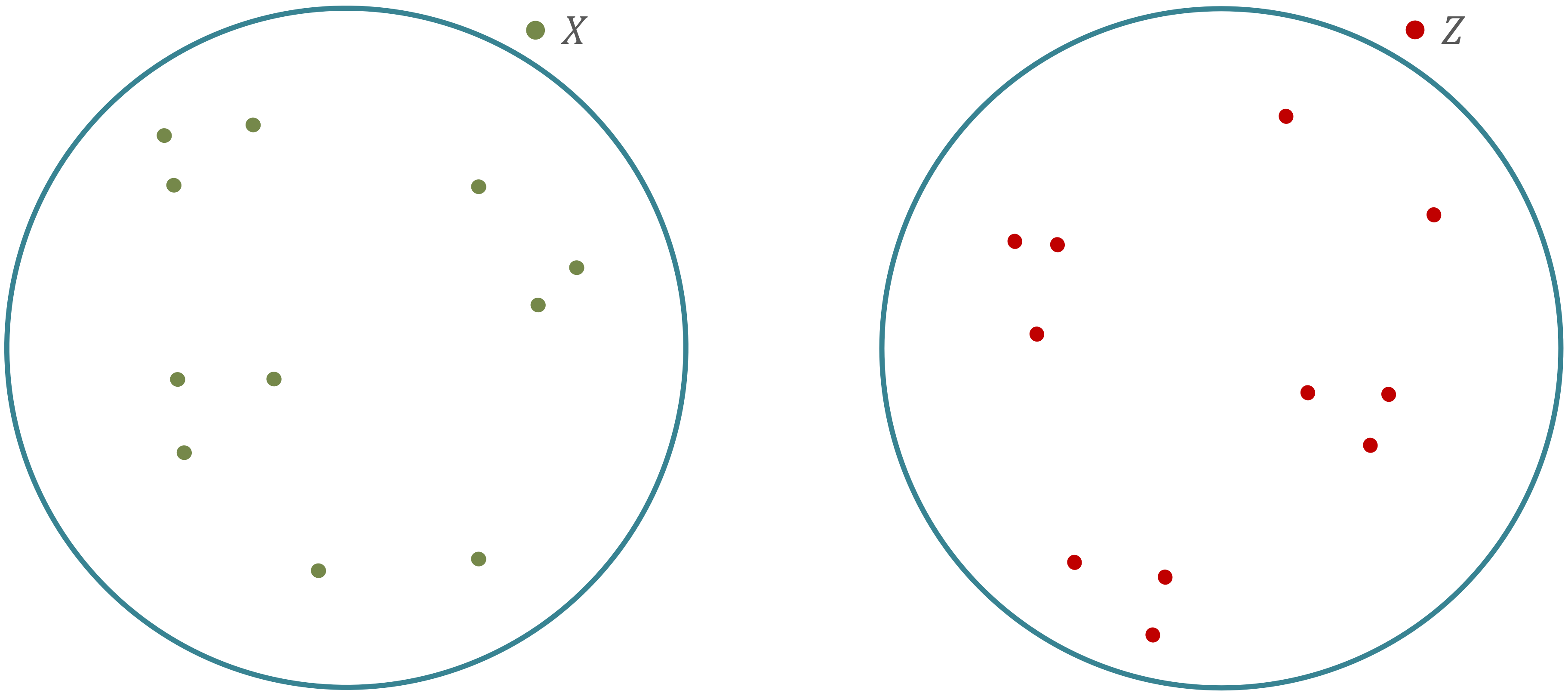
$$\underset{W\in O(d)}{\arg\min} \sum_i \|X_{i*}W - Z_{i*}\|$$

$$\begin{array}{r} \text{Txakur} \\ \text{Sagar} \\ \vdots \\ \text{Egutegi} \end{array} \begin{bmatrix} X_{1,*} \\ X_{2,*} \\ \vdots \\ X_{n,*} \end{bmatrix} \begin{bmatrix} W \end{bmatrix} \approx \begin{bmatrix} Z_{1,*} \\ Z_{2,*} \\ \vdots \\ Z_{n,*} \end{bmatrix} \begin{array}{l} \text{Dog} \\ \text{Apple} \\ \vdots \\ \text{Calendar} \end{array}$$

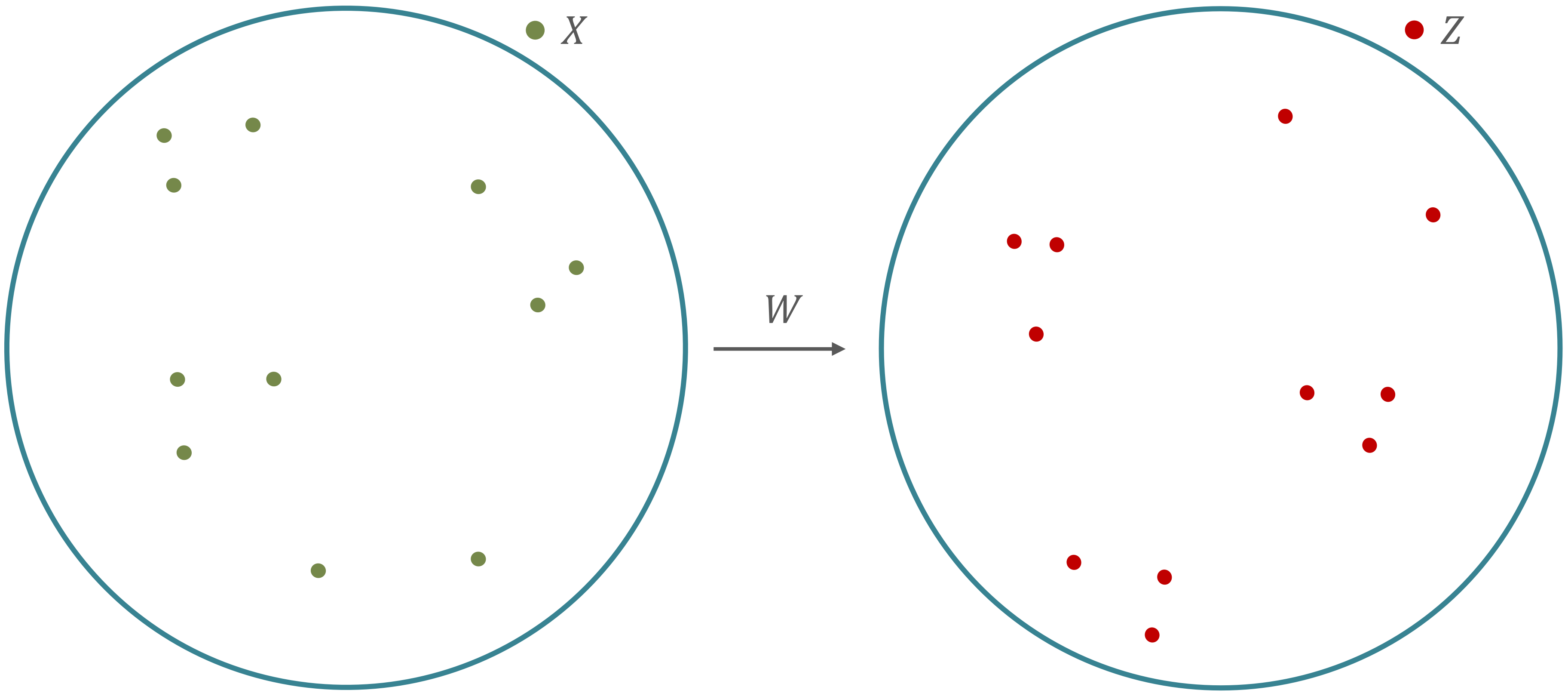
Cross-lingual embedding mappings



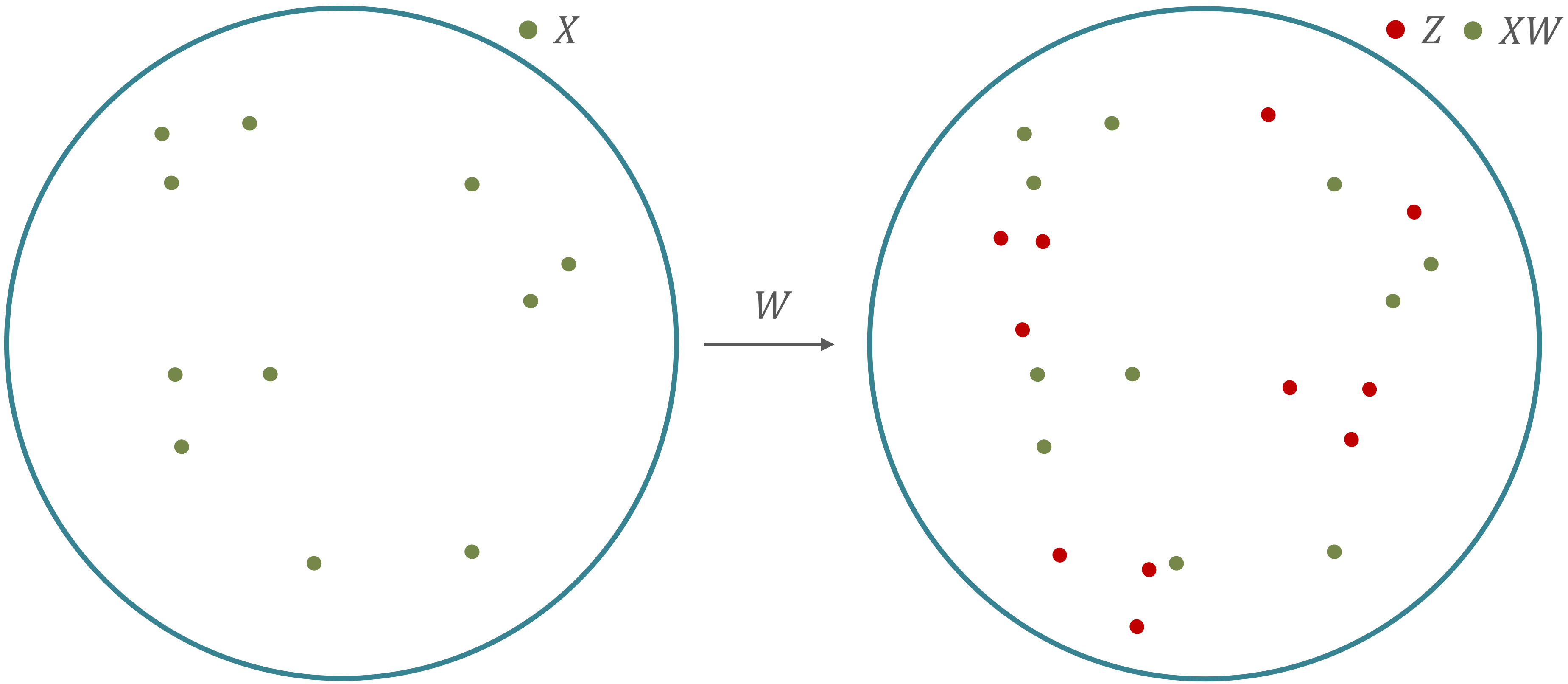
Cross-lingual embedding mappings



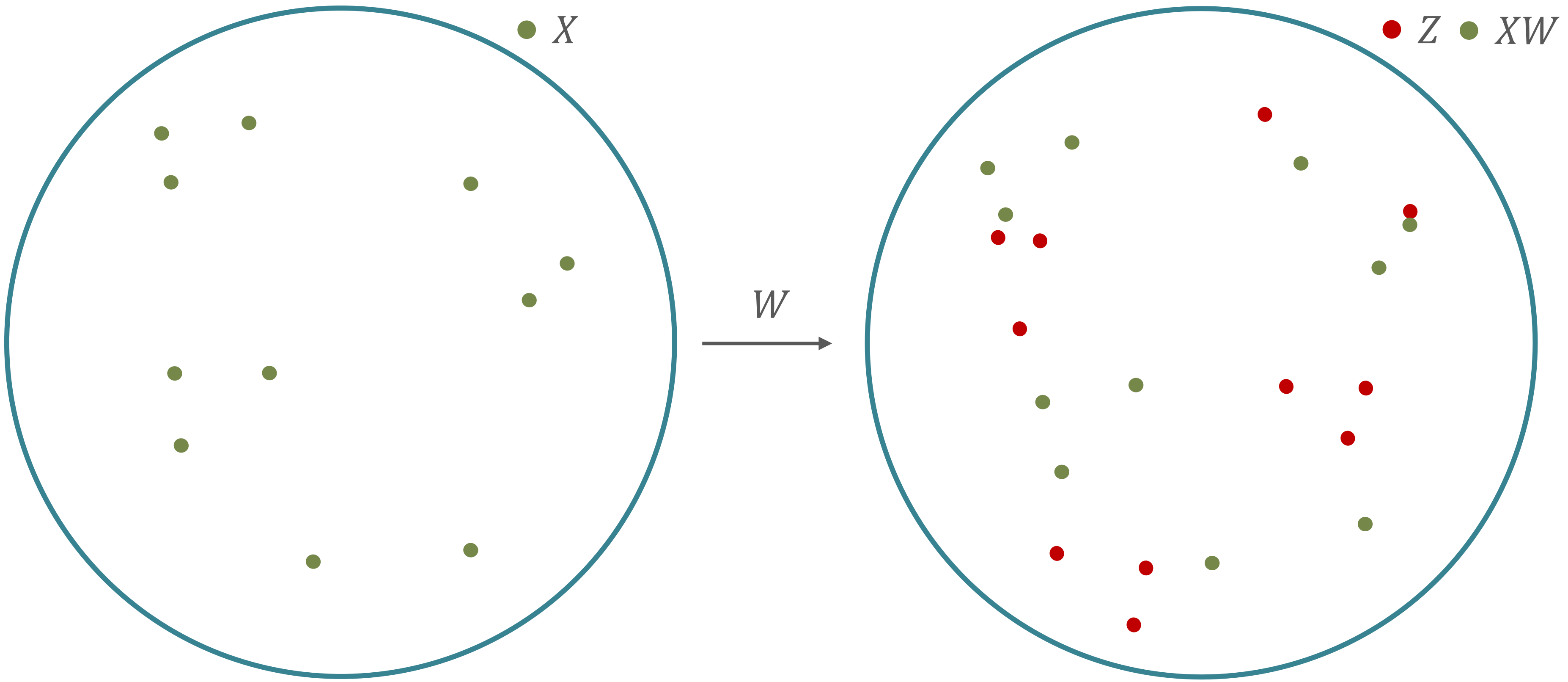
Cross-lingual embedding mappings



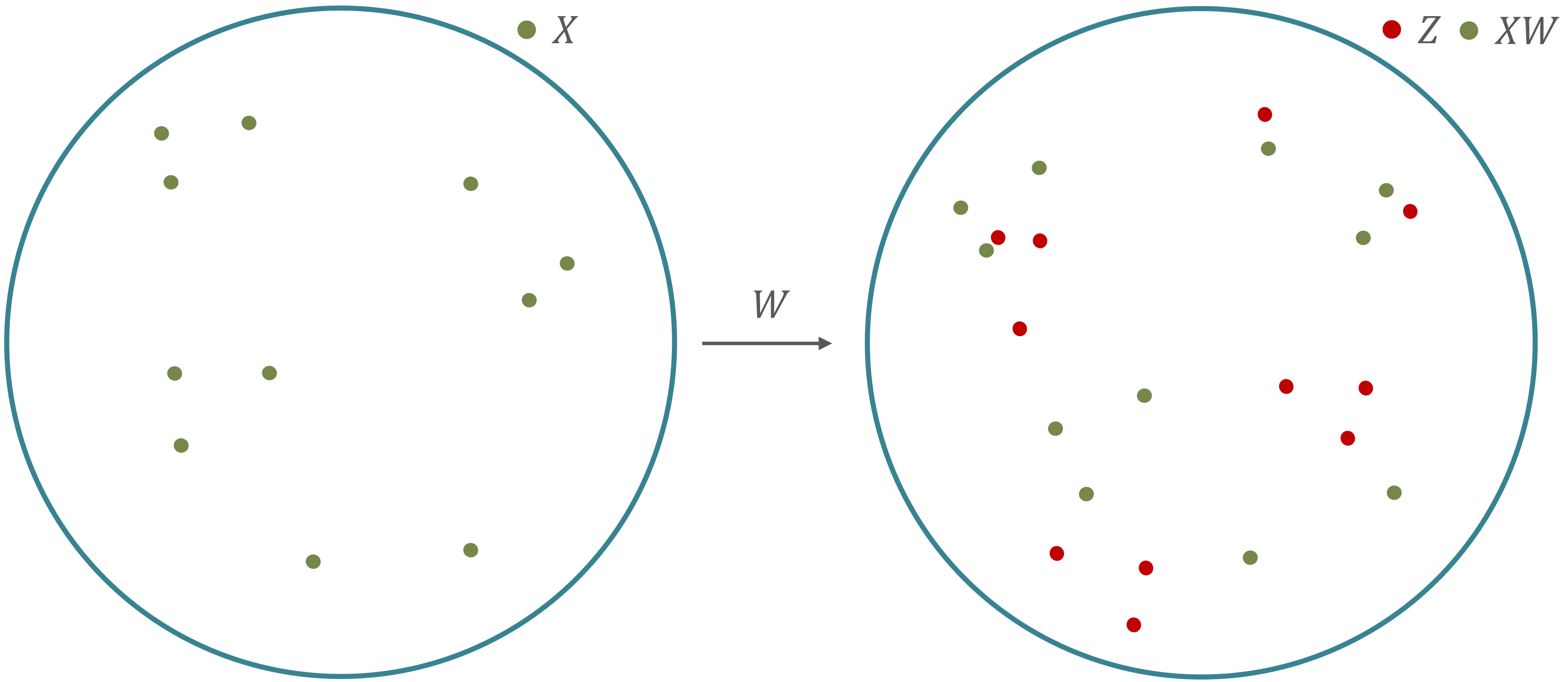
Cross-lingual embedding mappings



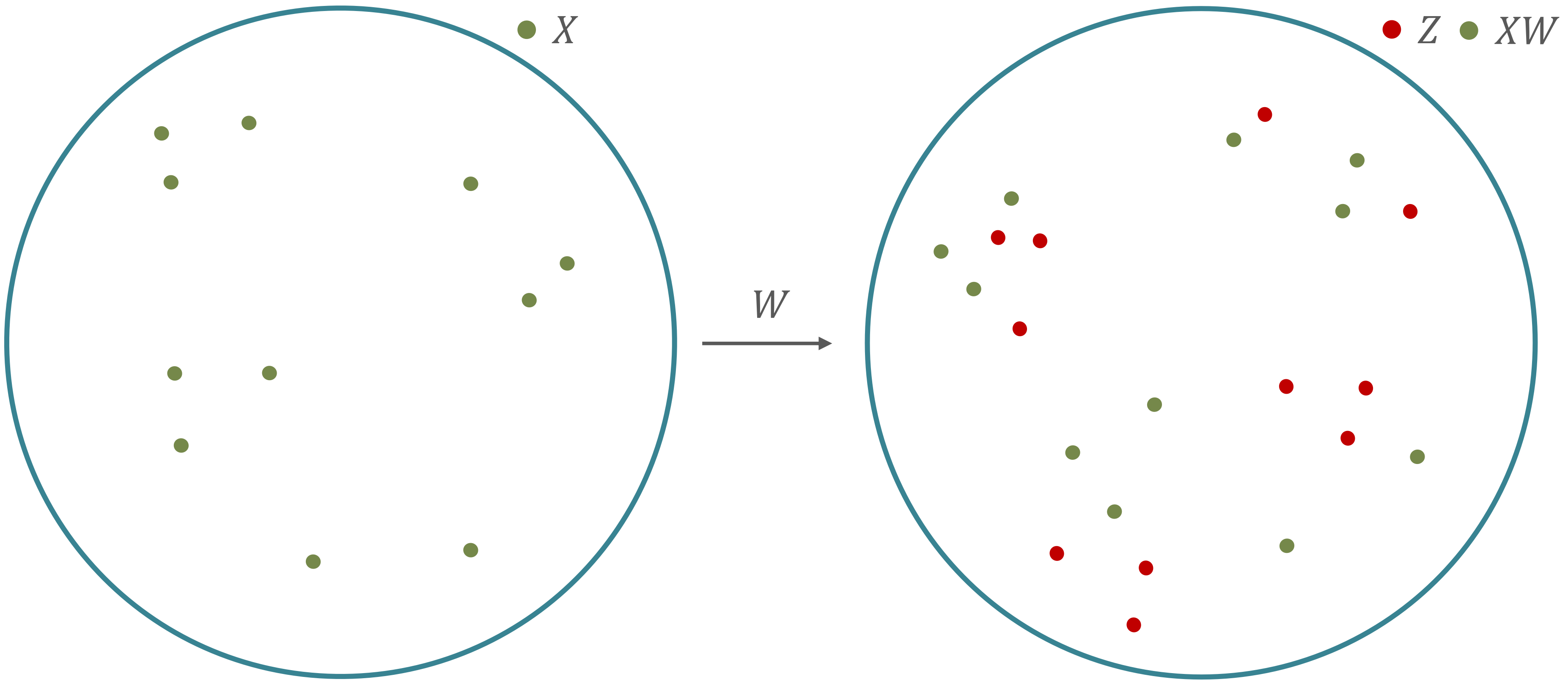
Cross-lingual embedding mappings



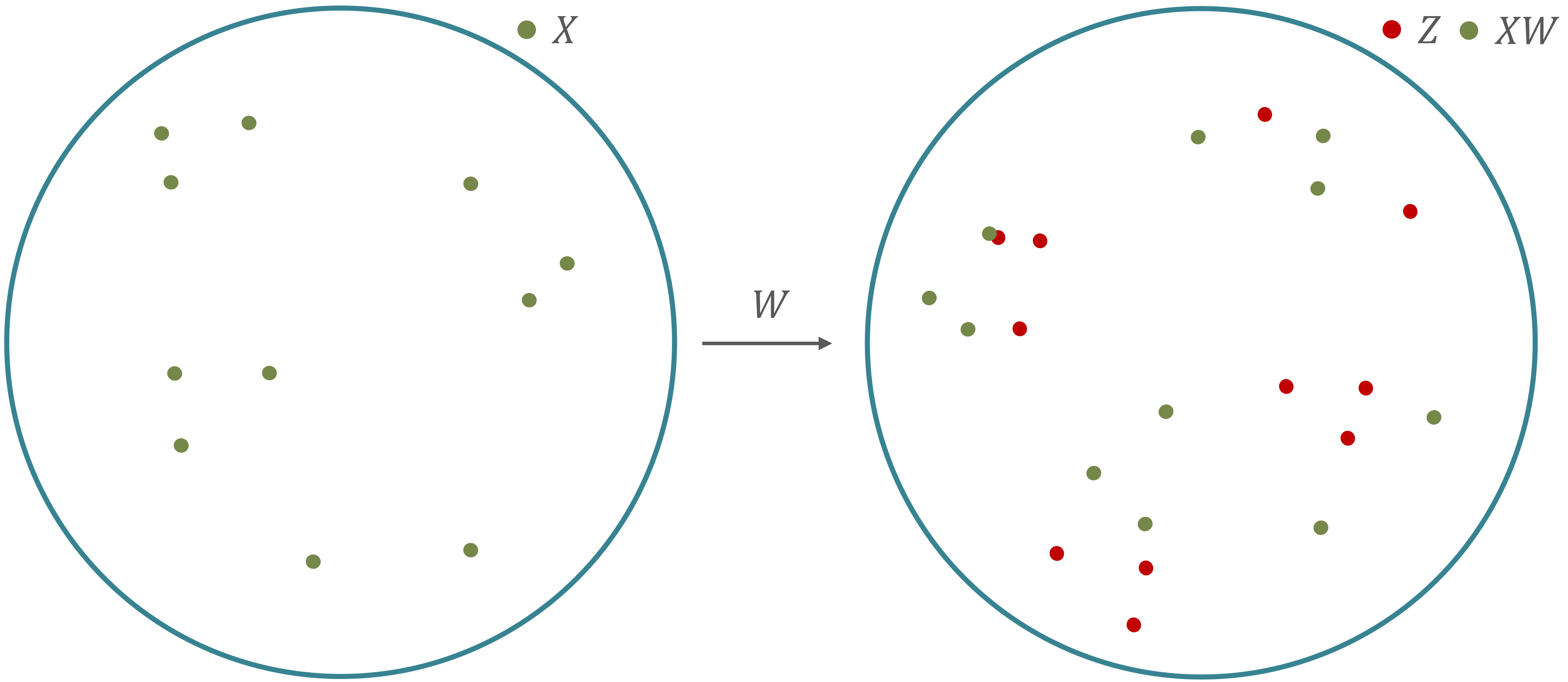
Cross-lingual embedding mappings



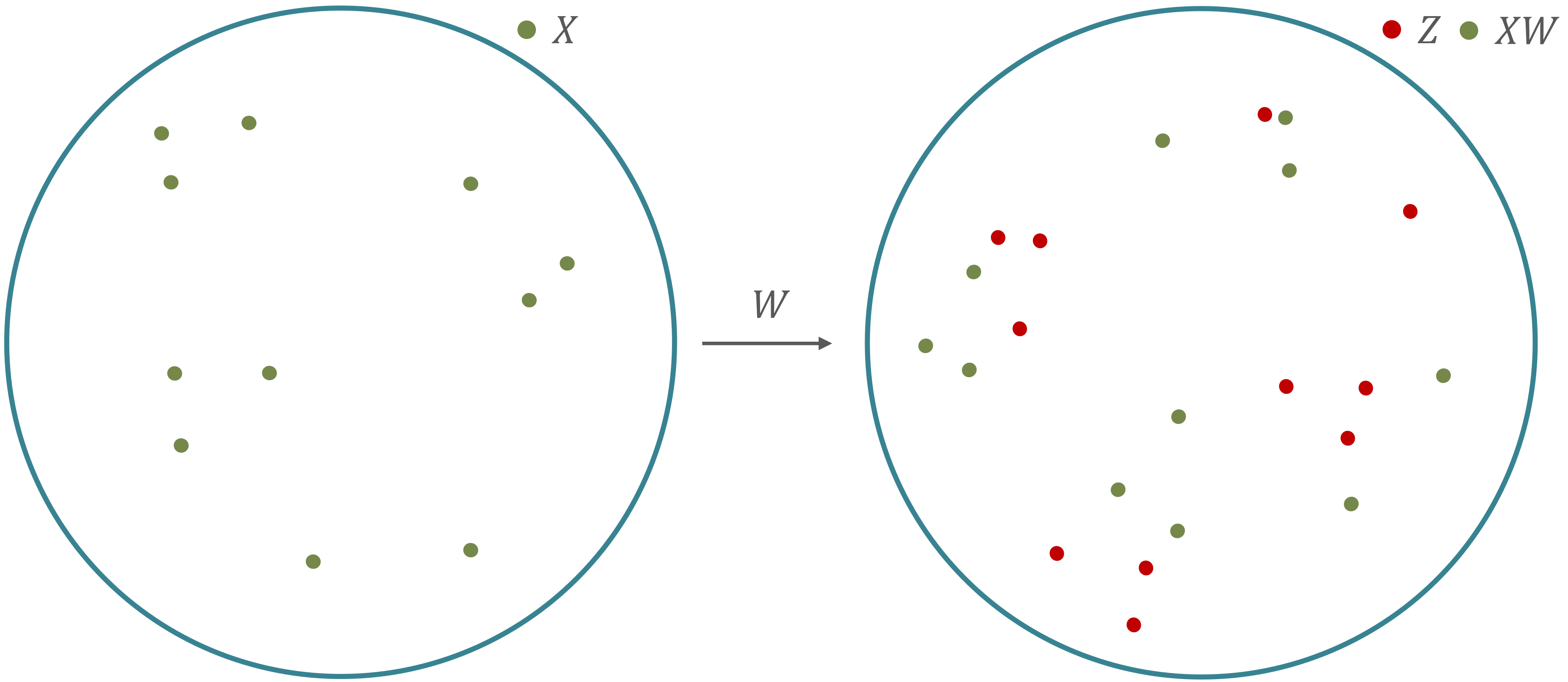
Cross-lingual embedding mappings



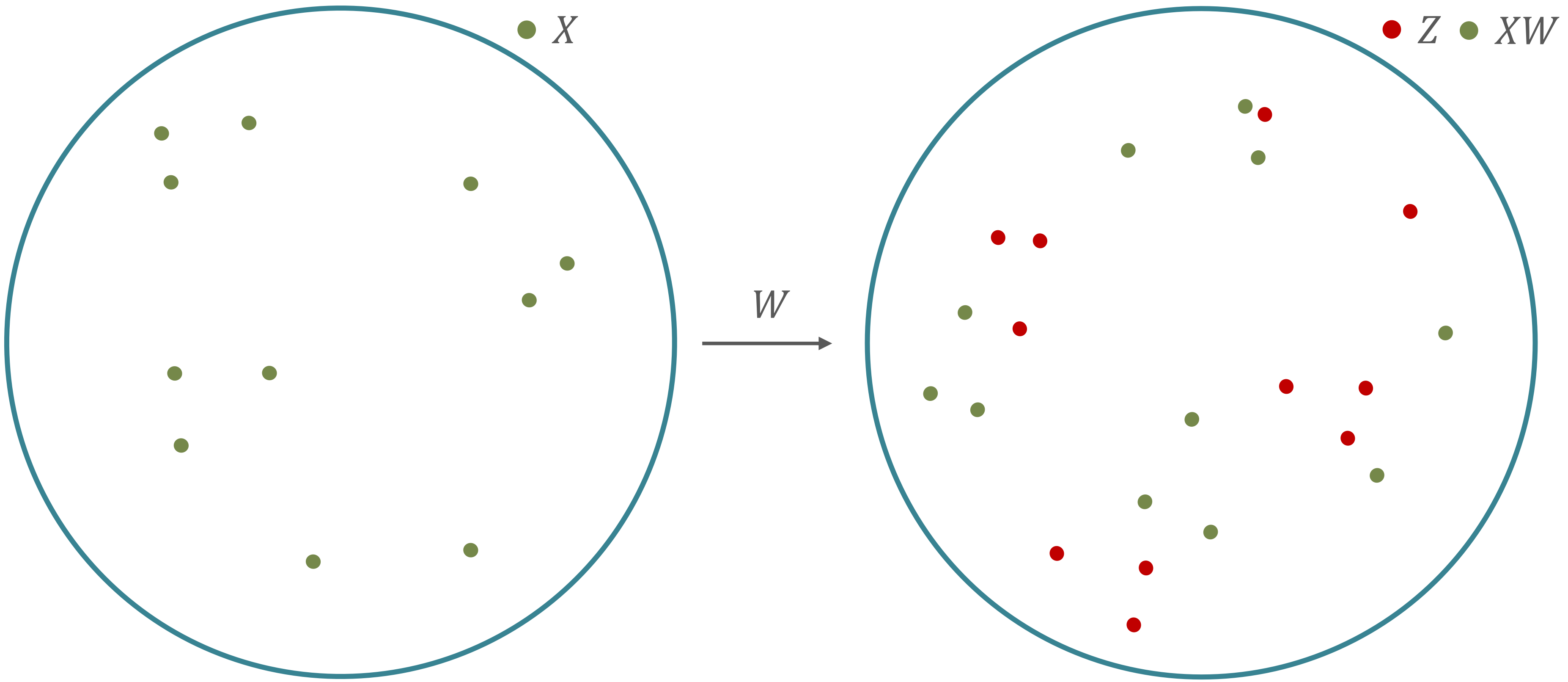
Cross-lingual embedding mappings



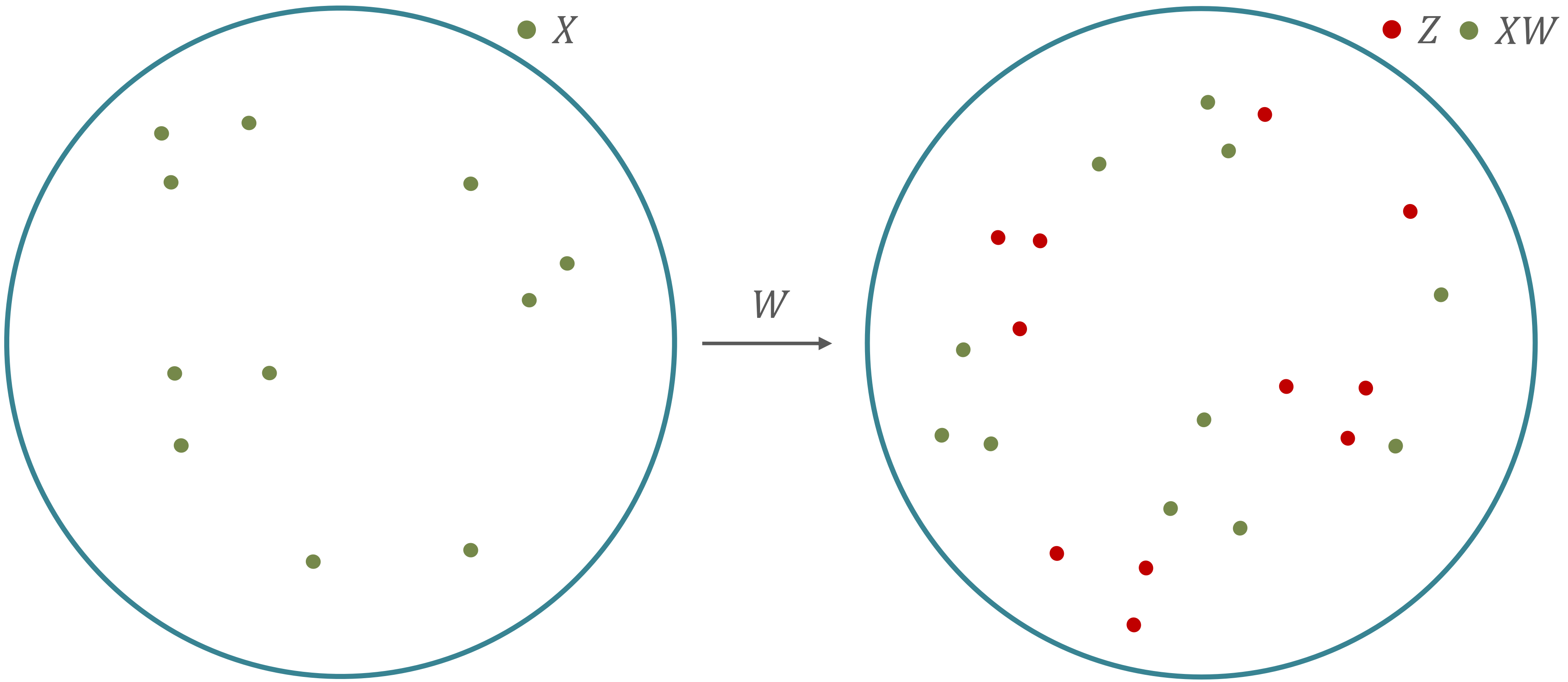
Cross-lingual embedding mappings



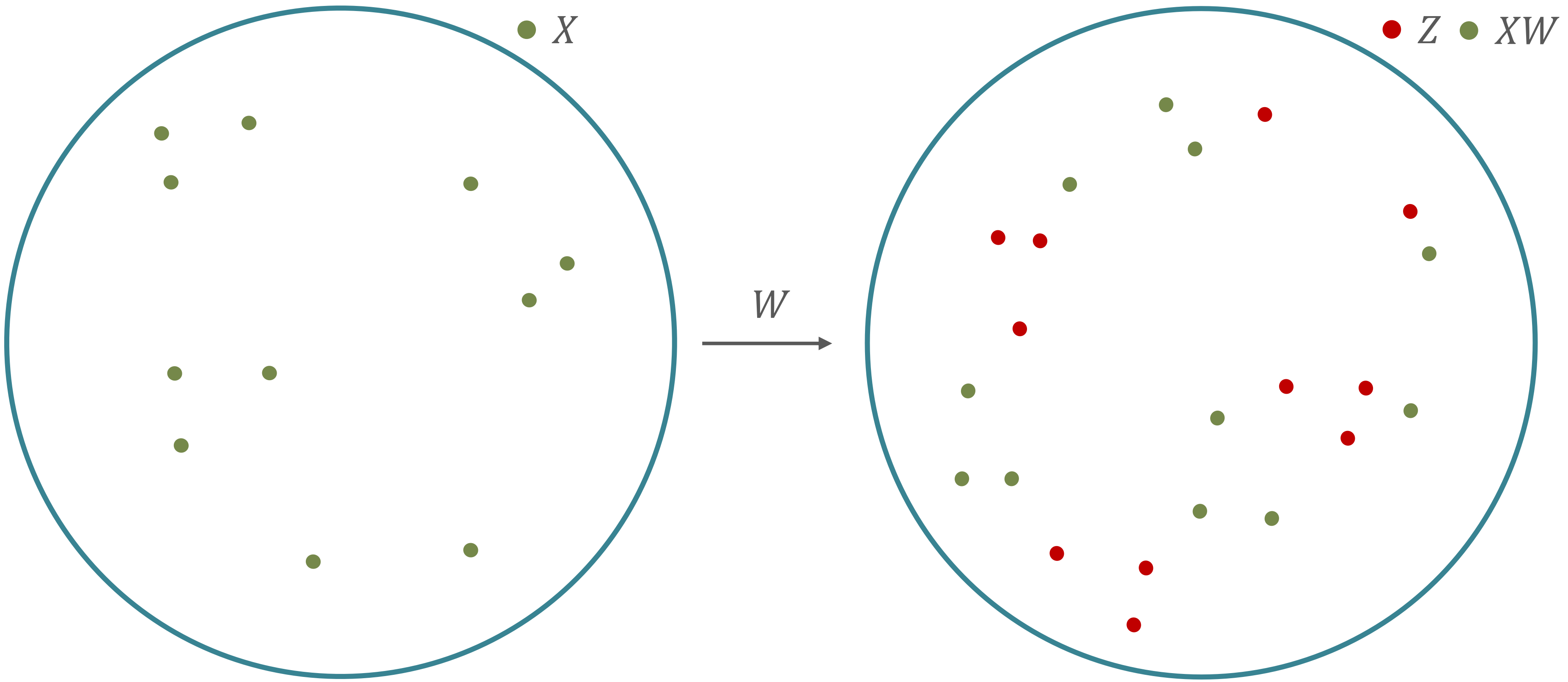
Cross-lingual embedding mappings



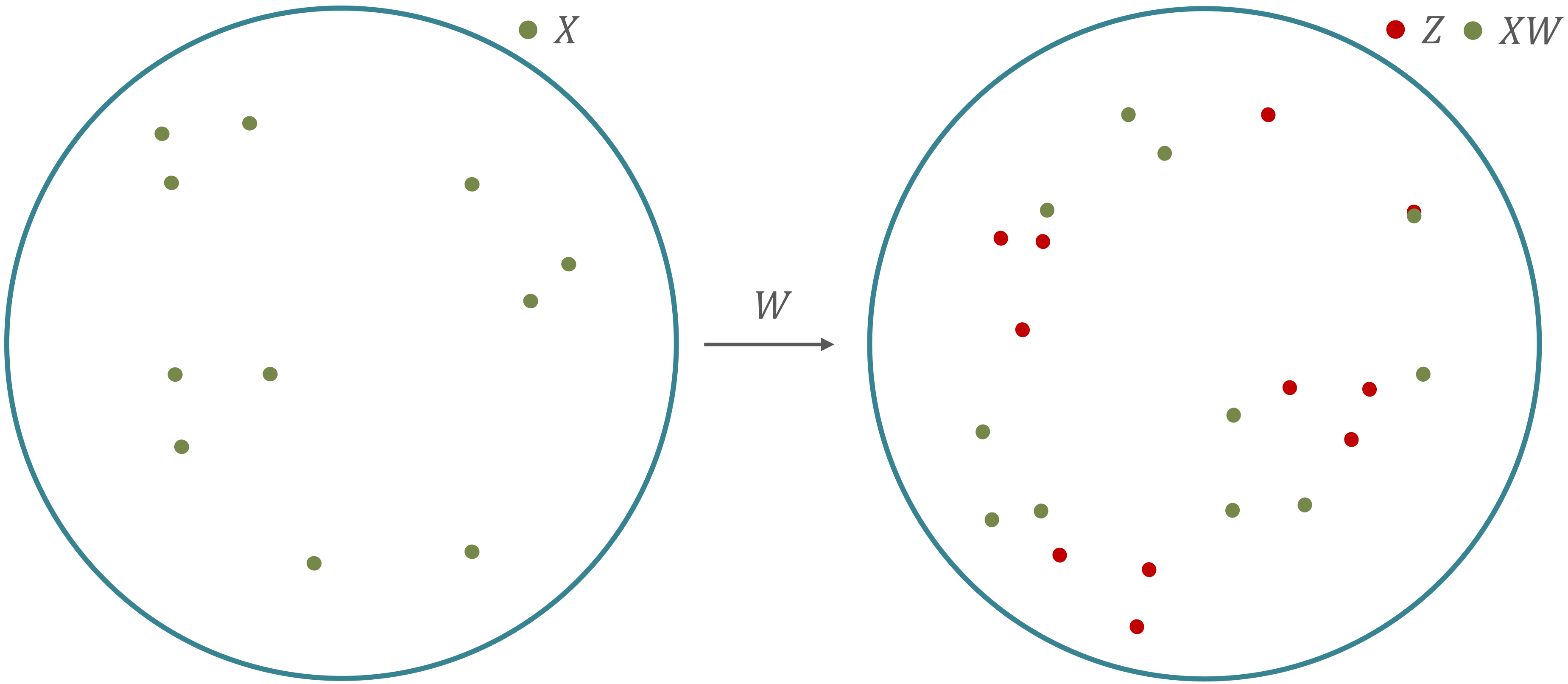
Cross-lingual embedding mappings



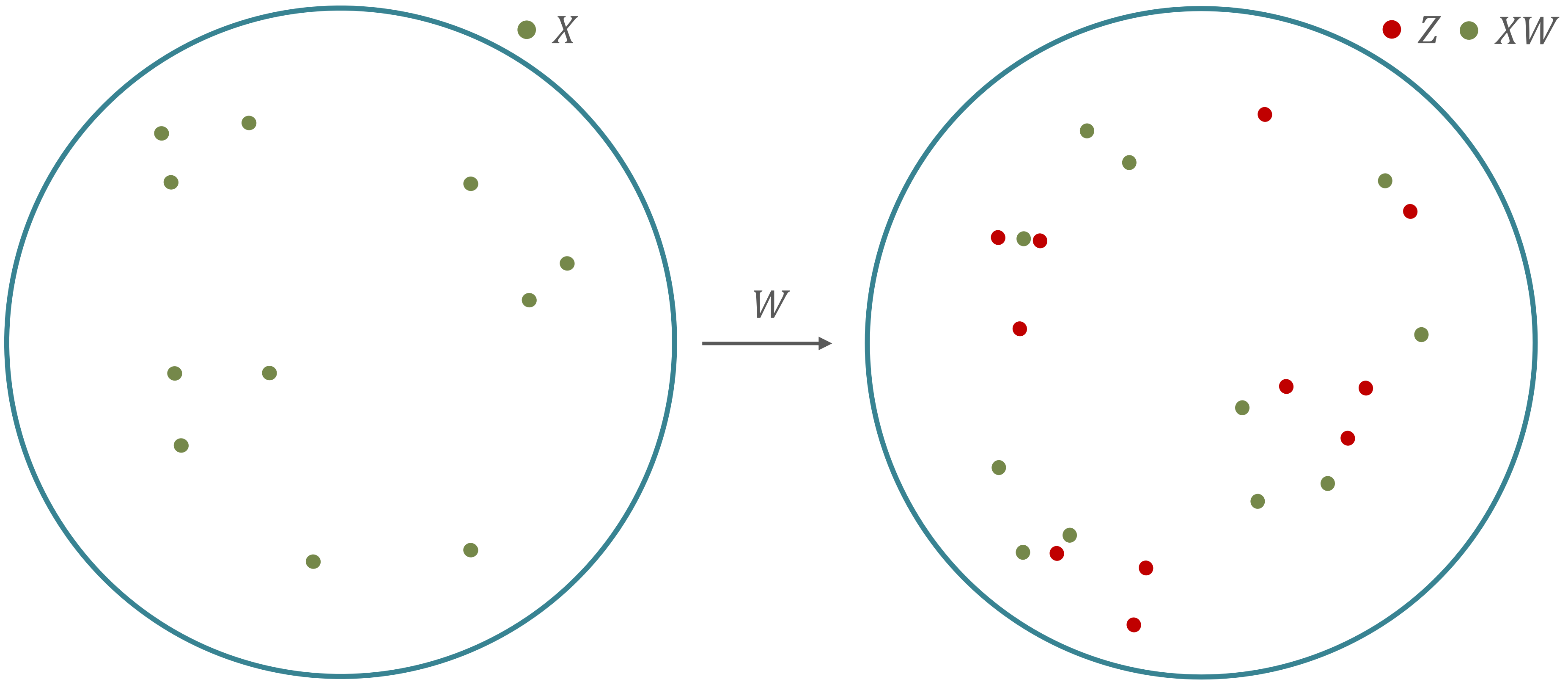
Cross-lingual embedding mappings



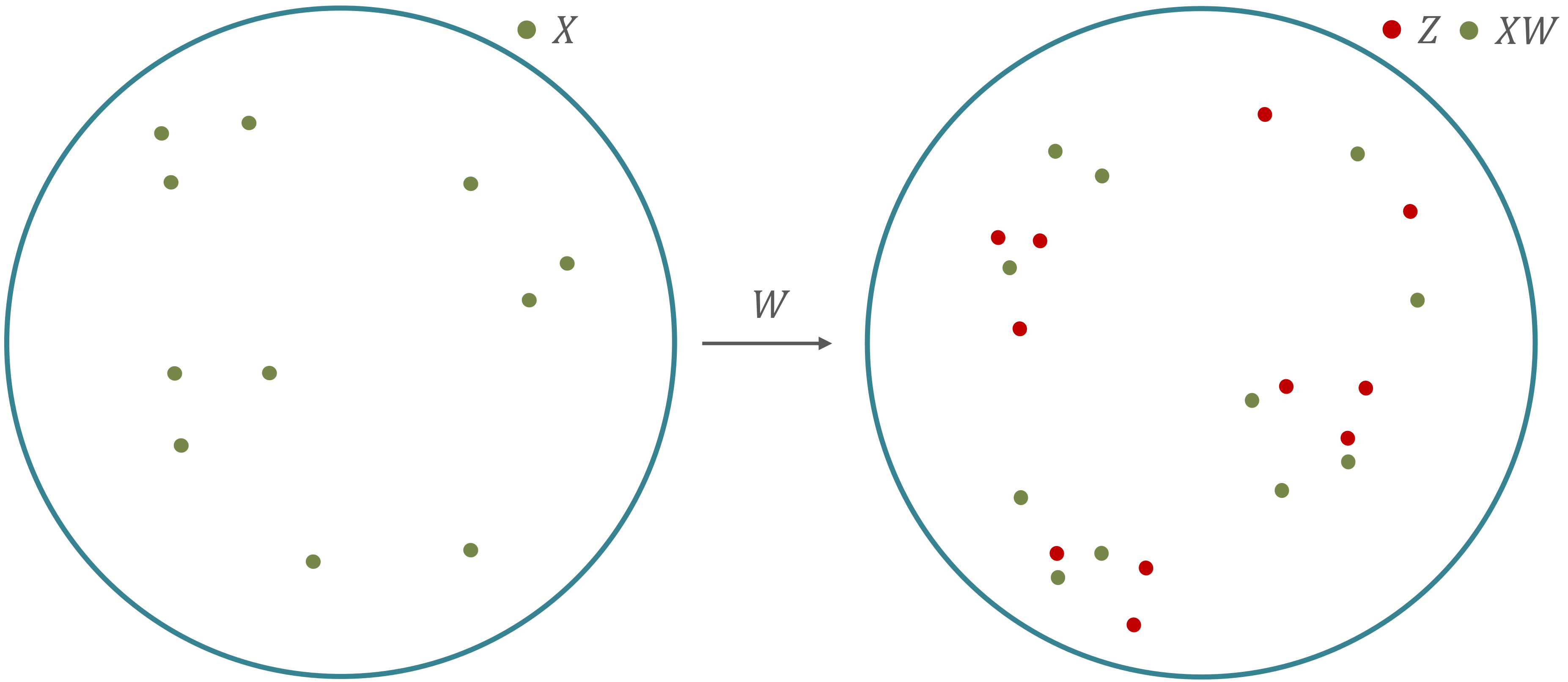
Cross-lingual embedding mappings



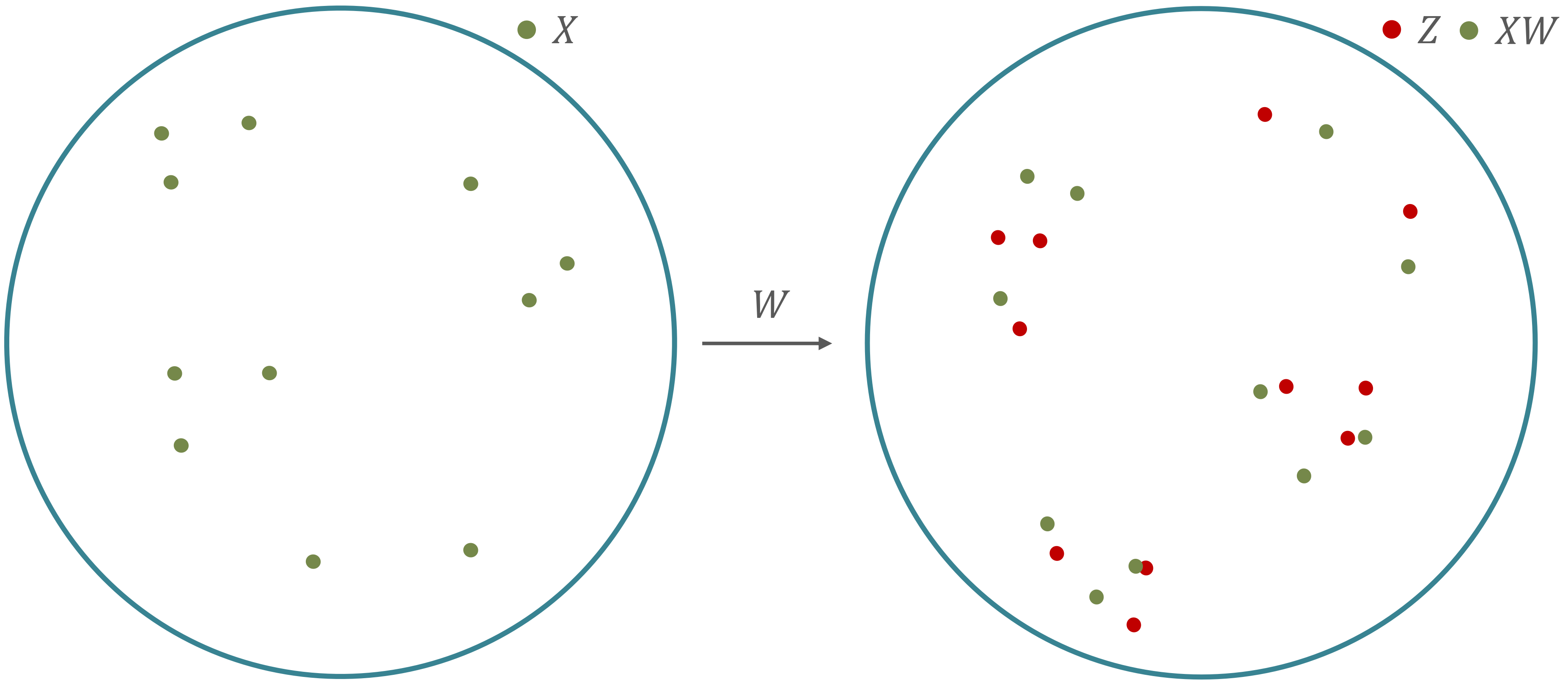
Cross-lingual embedding mappings



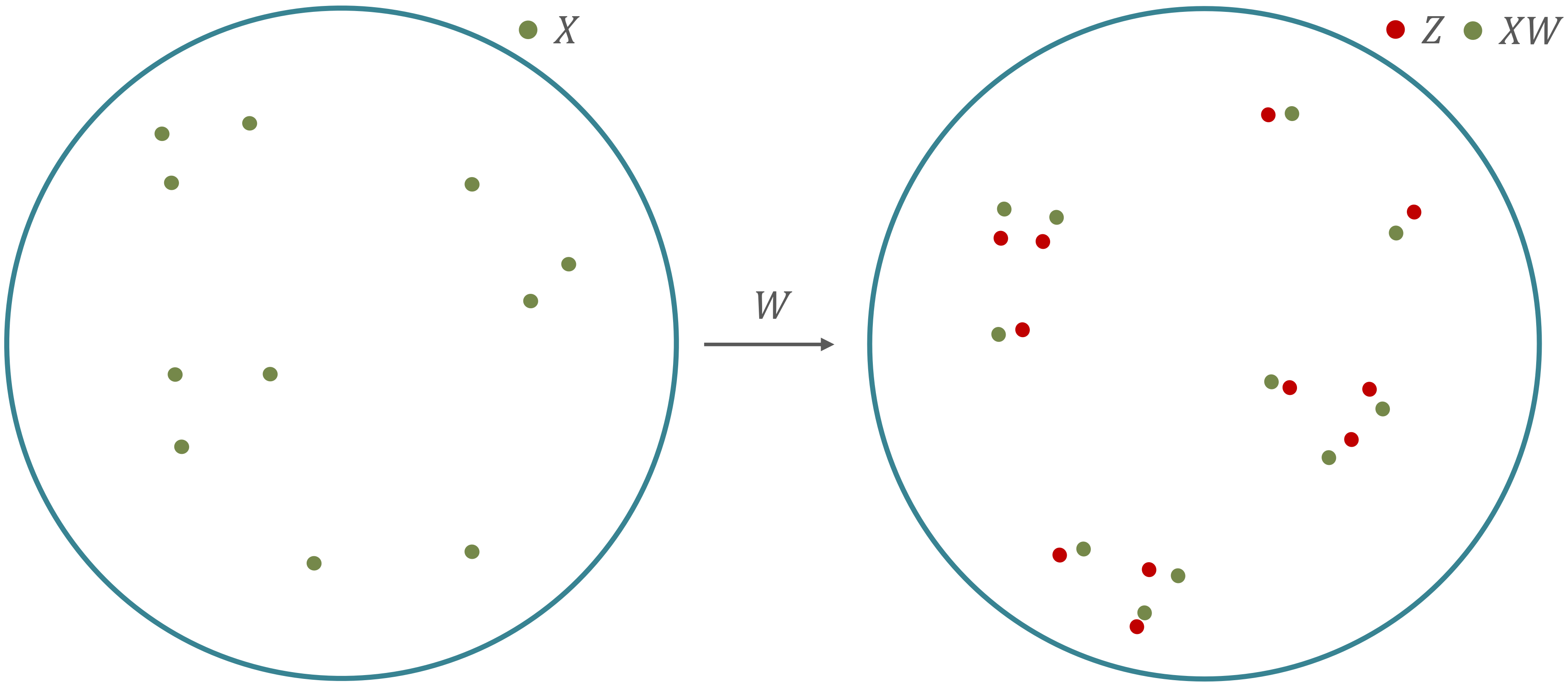
Cross-lingual embedding mappings



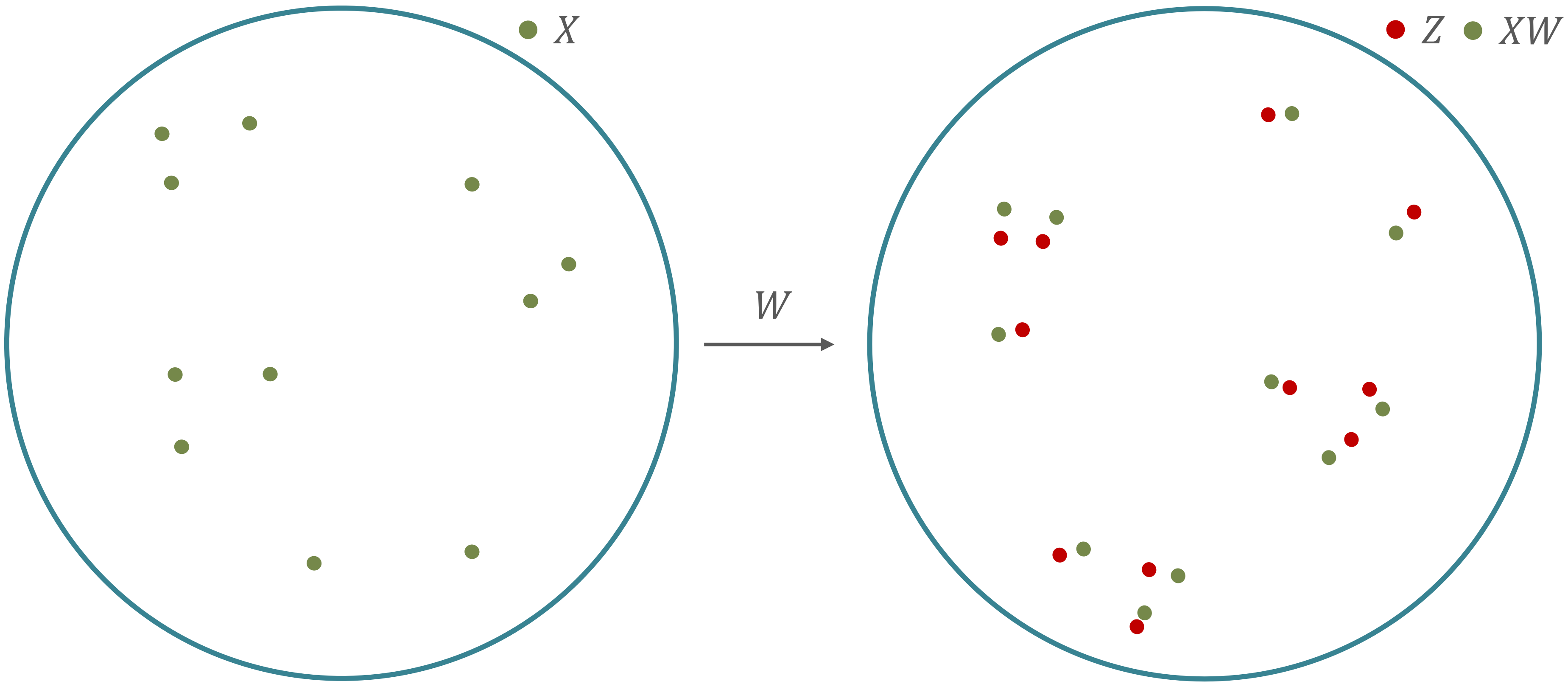
Cross-lingual embedding mappings



Cross-lingual embedding mappings



Cross-lingual embedding mappings



$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*}W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*}W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL'17)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*}W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL'17)

Dictionary

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*}W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL'17)

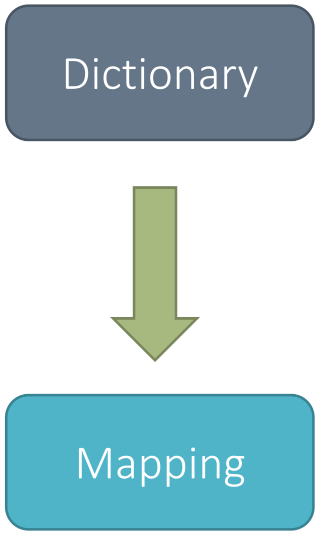
Dictionary

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2$$

Cross-lingual embedding mappings

<u>Self-learning</u>

(Artetxe et al., ACL'17)

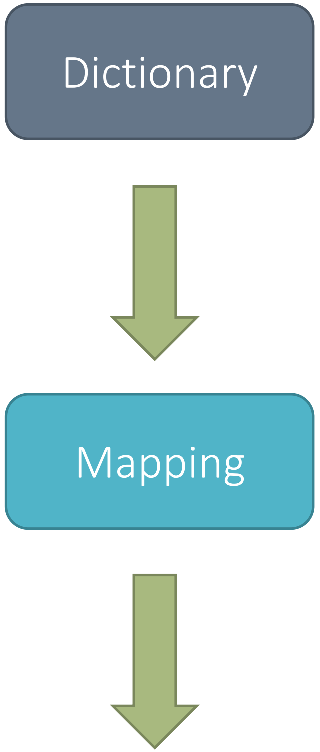


$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL'17)

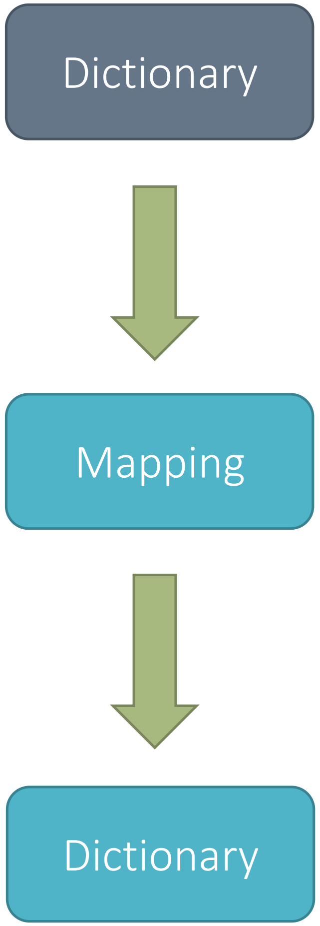


$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL'17)



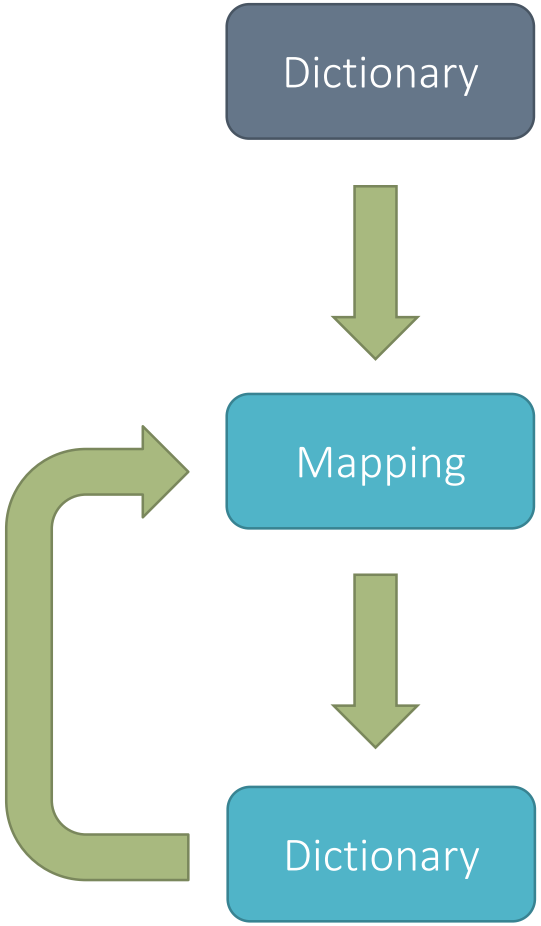
$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL'17)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

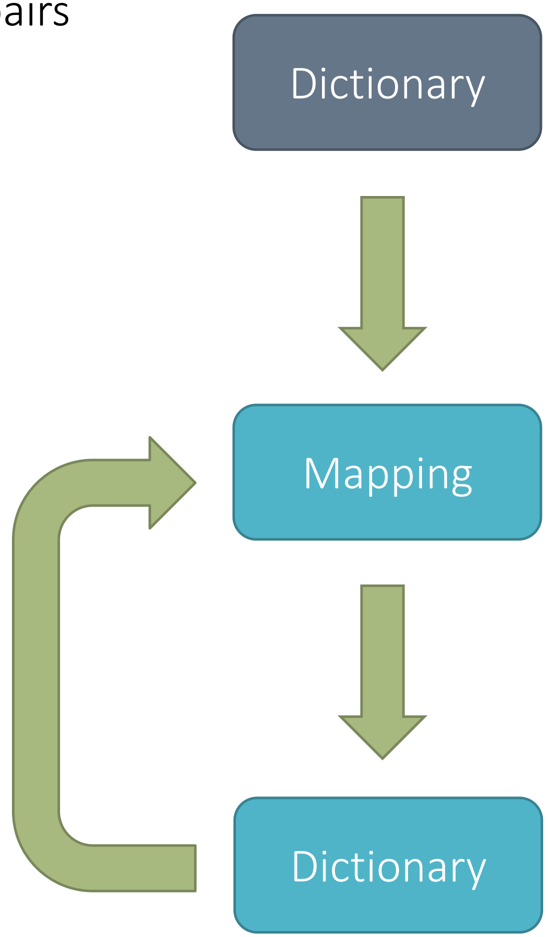


Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL'17)

- 25 word pairs



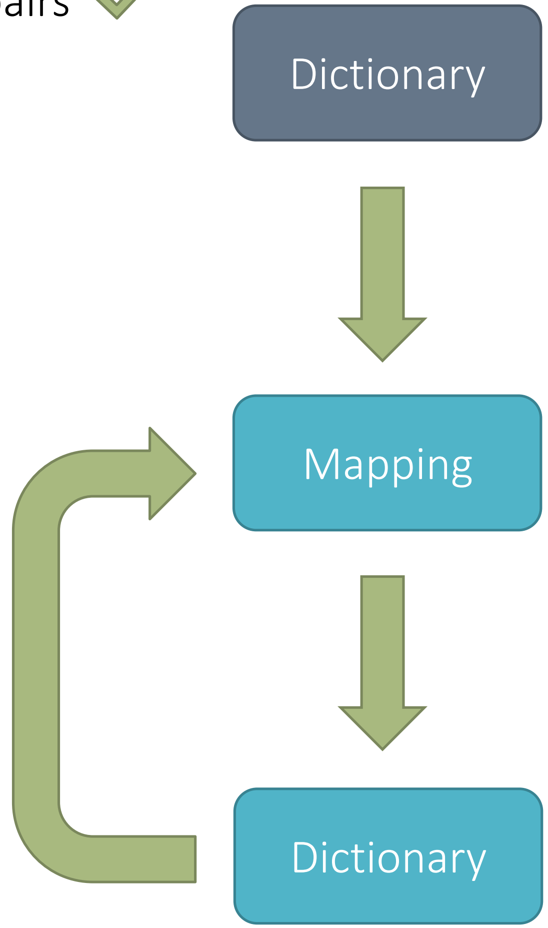
$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*}W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL’17)

- 25 word pairs ✓



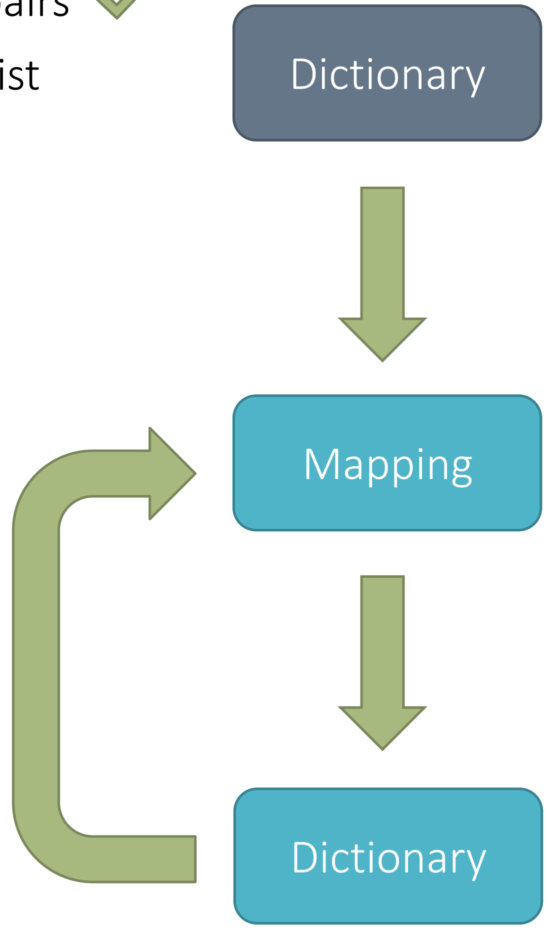
$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2$$

Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL'17)

- 25 word pairs ✓
- Numeral list



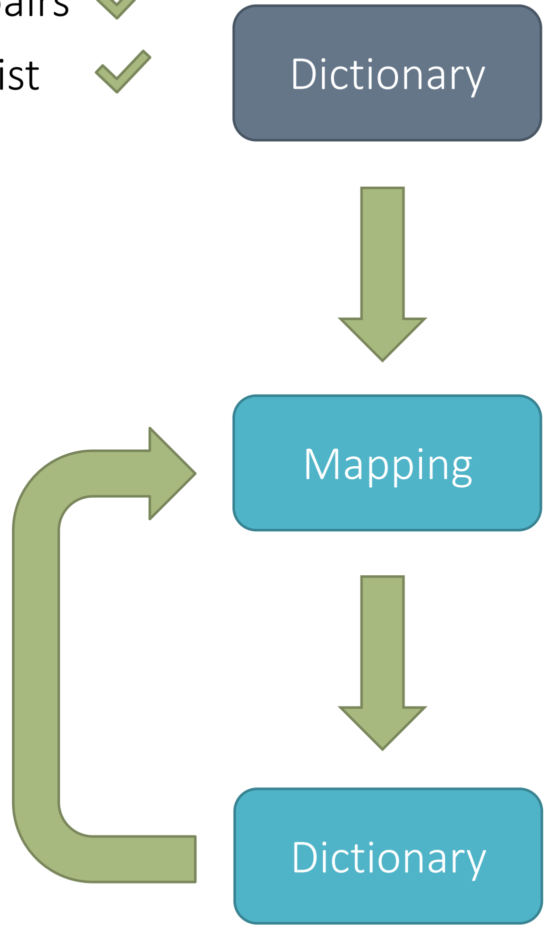
$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL’17)

- 25 word pairs ✓
- Numeral list ✓



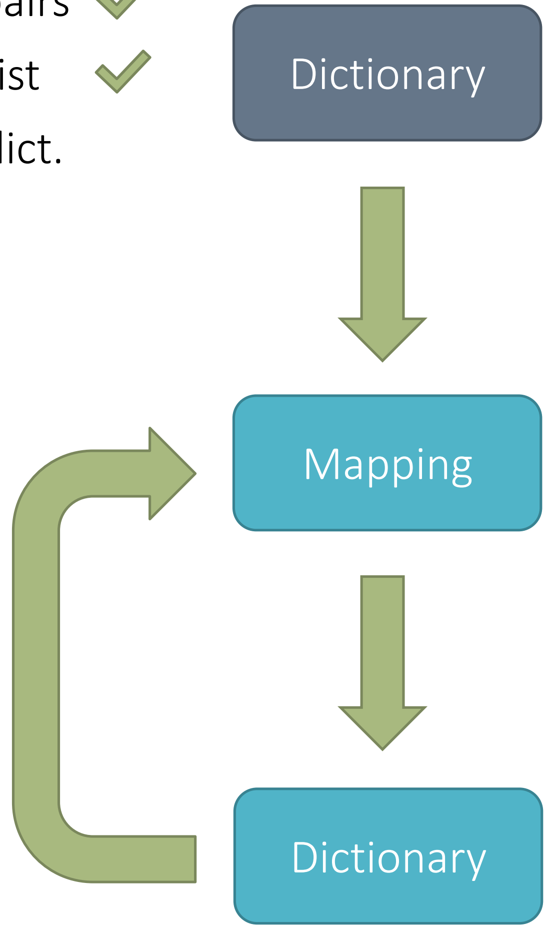
$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*}W - Z_{j*} \right\|^2$$

Cross-lingual embedding mappings

Self-learning

(Artetxe et al., ACL'17)

- 25 word pairs ✓
- Numeral list ✓
- Random dict.



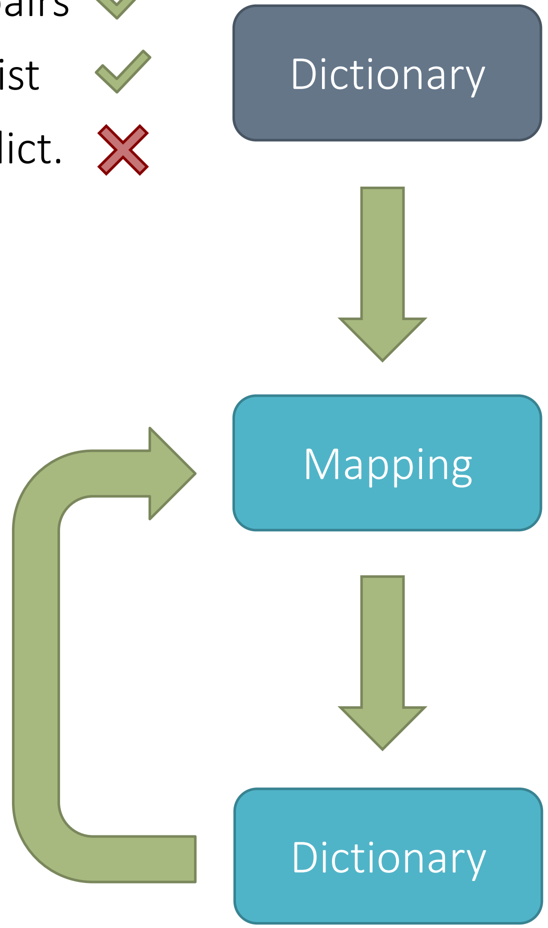
$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2$$

Cross-lingual embedding mappings

Self-learning

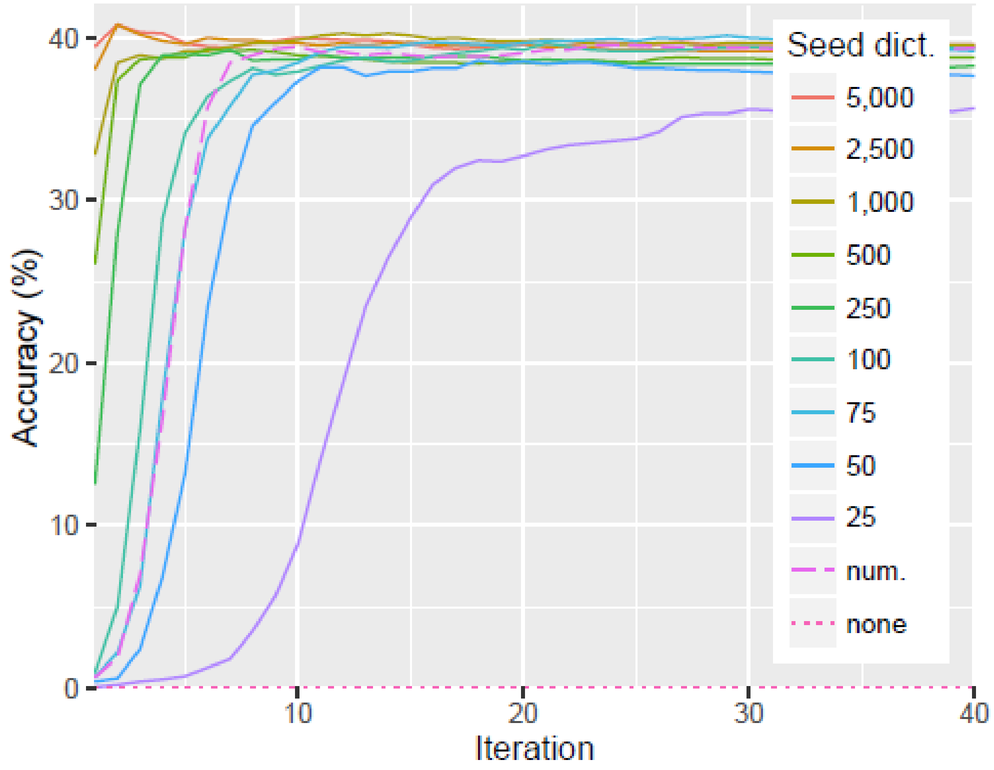
(Artetxe et al., ACL'17)

- 25 word pairs ✓
- Numeral list ✓
- Random dict. ✗



$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2$$

Cross-lingual embedding mappings

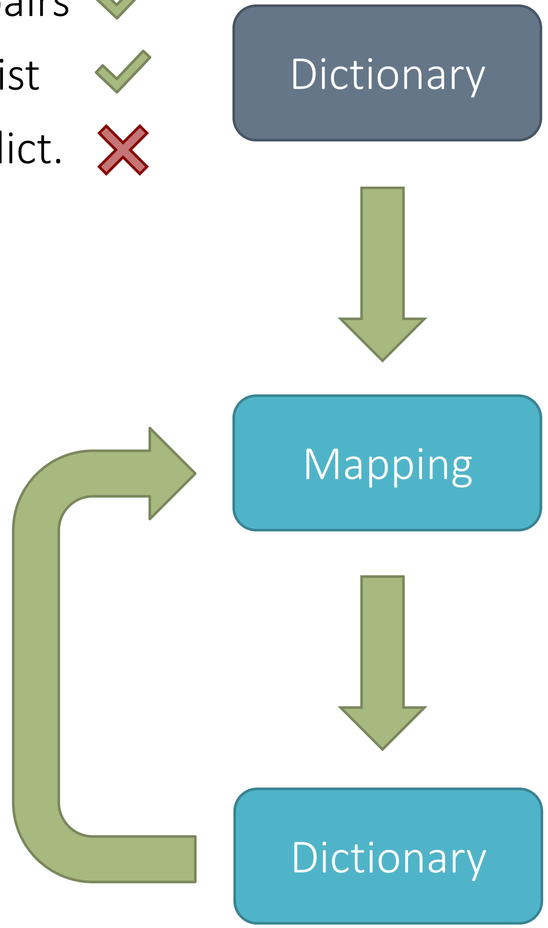


$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2$$

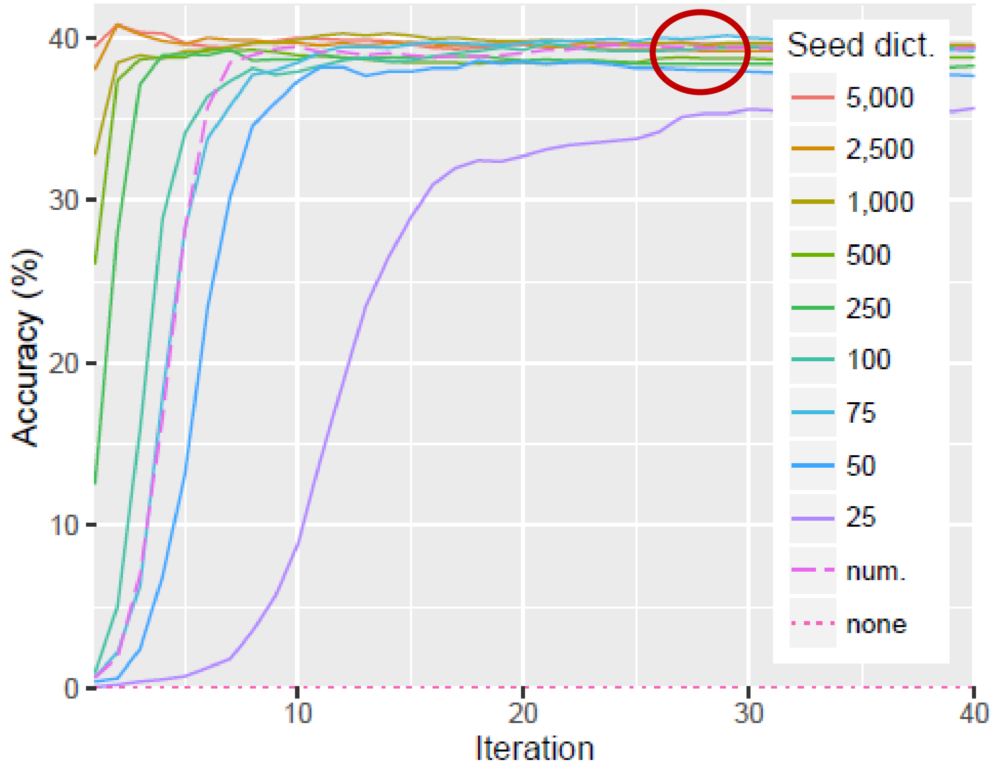
Self-learning

(Artetxe et al., ACL'17)

- 25 word pairs ✓
- Numeral list ✓
- Random dict. ✗



Cross-lingual embedding mappings

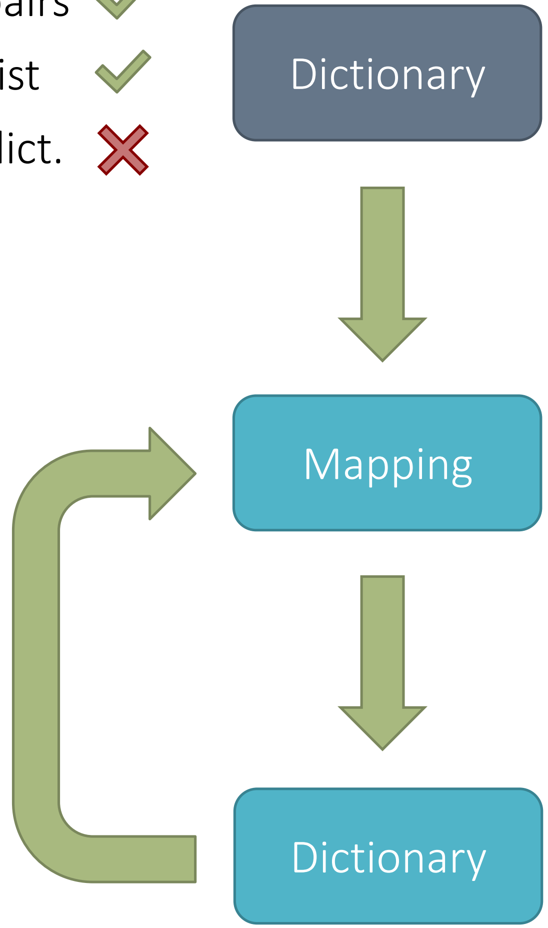


$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$

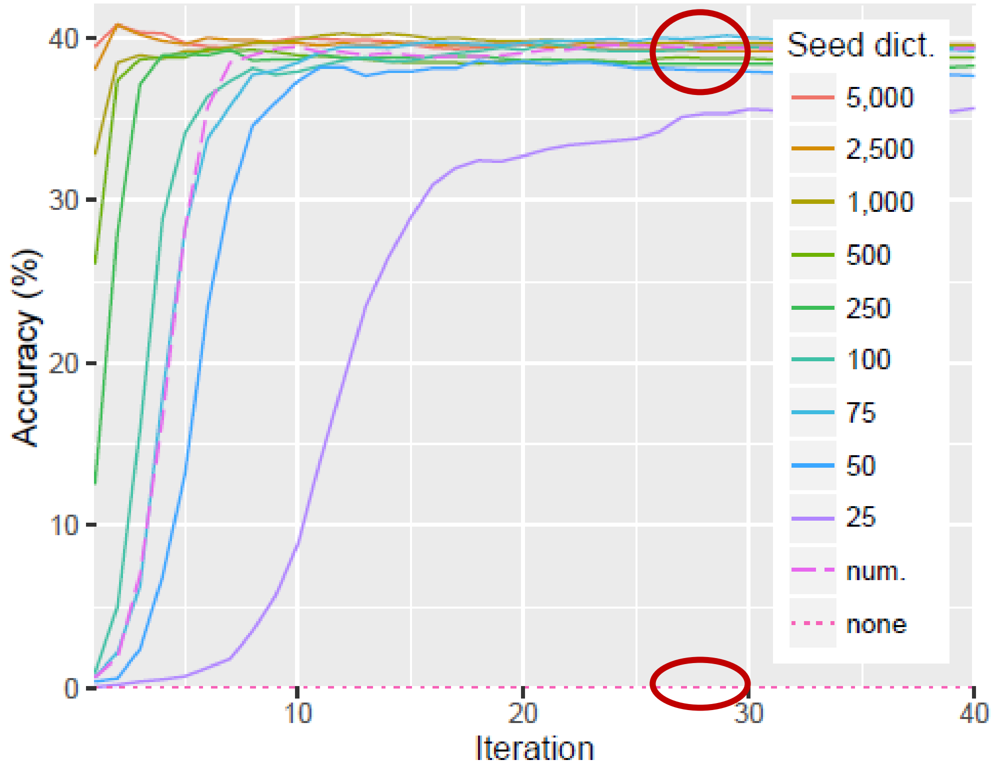
Self-learning

(Artetxe et al., ACL'17)

- 25 word pairs ✓
- Numeral list ✓
- Random dict. ✗



Cross-lingual embedding mappings

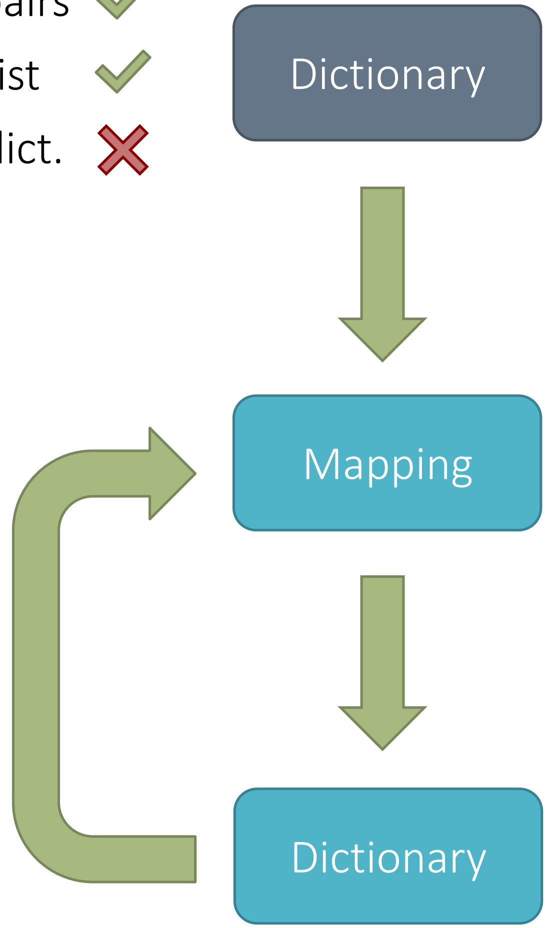


$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2$$

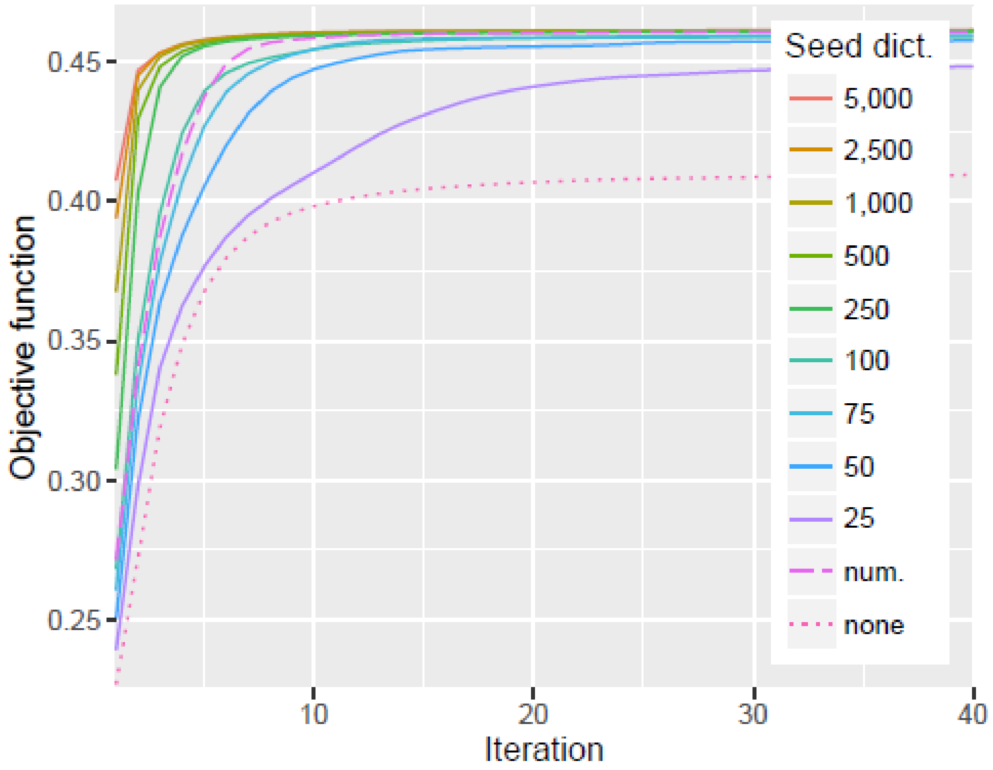
Self-learning

(Artetxe et al., ACL'17)

- 25 word pairs ✓
- Numeral list ✓
- Random dict. ✗



Cross-lingual embedding mappings

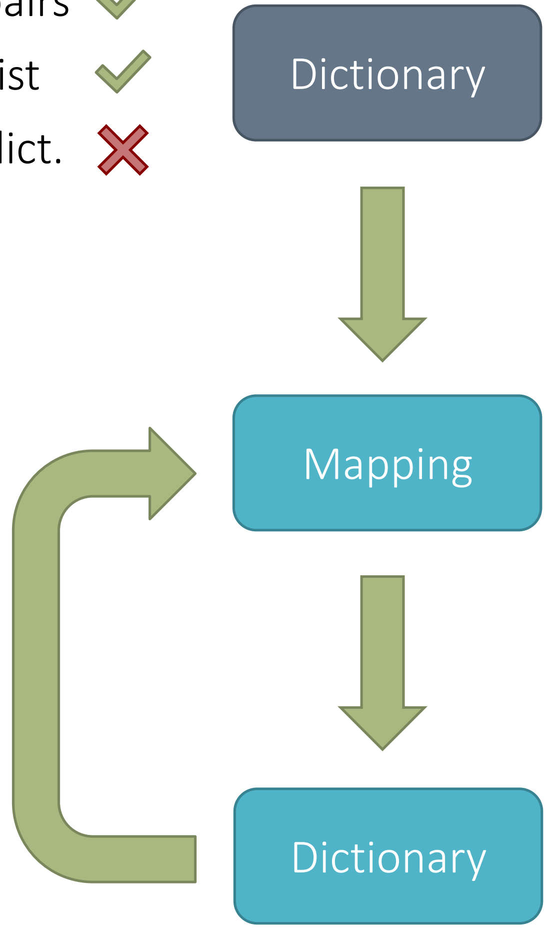


$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2$$

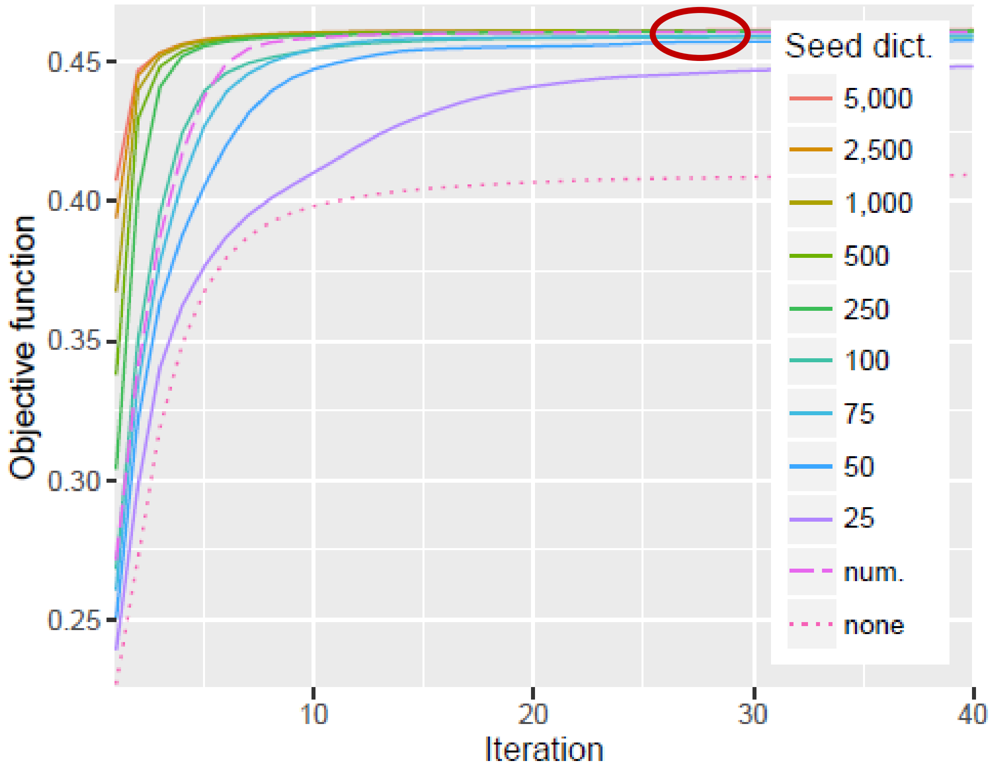
Self-learning

(Artetxe et al., ACL'17)

- 25 word pairs ✓
- Numeral list ✓
- Random dict. ✗



Cross-lingual embedding mappings

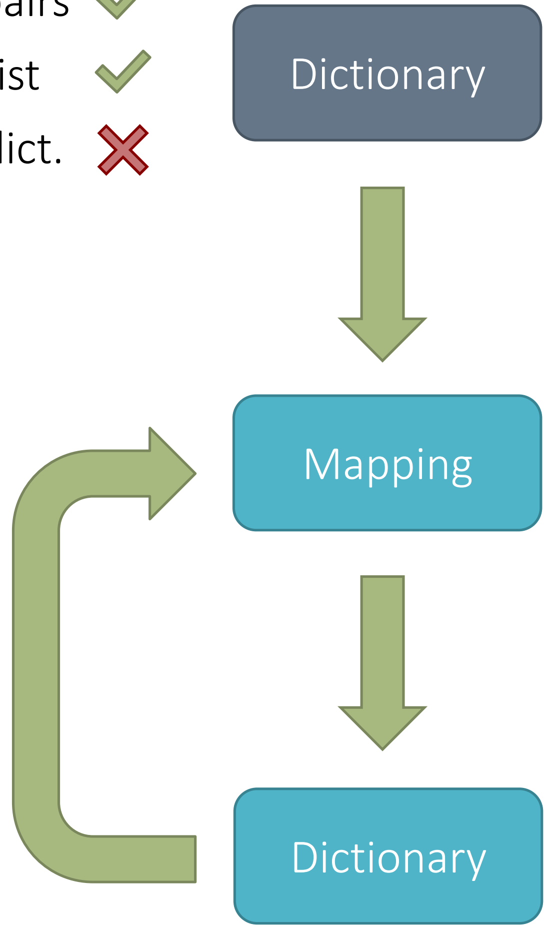


$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2$$

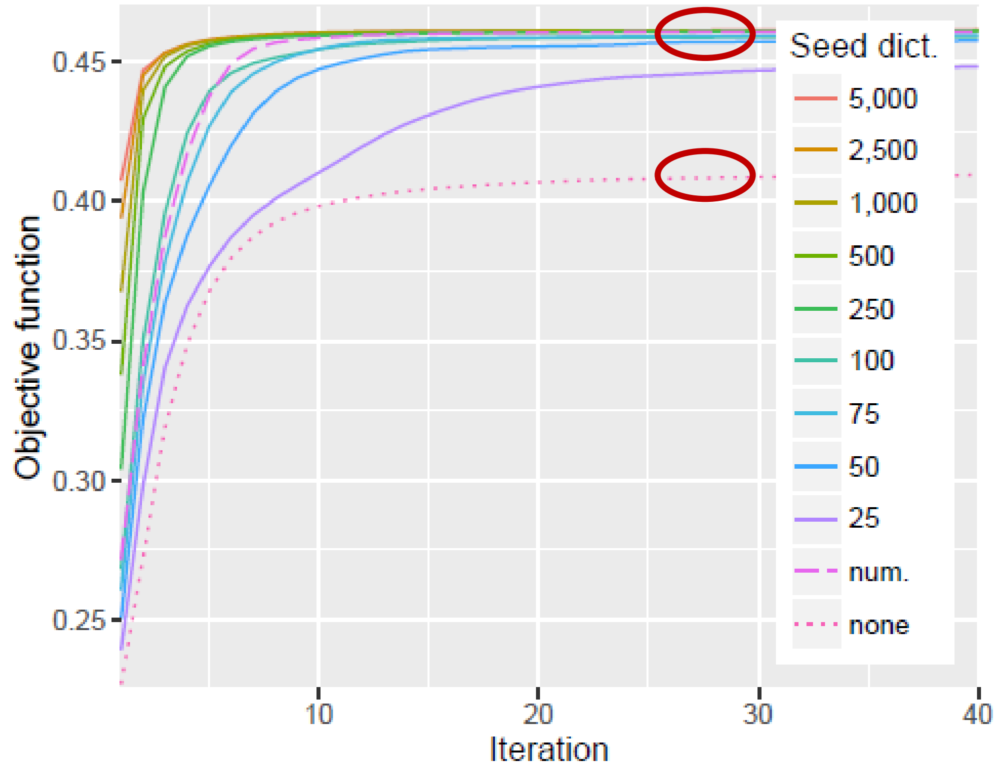
<u>Self-learning</u>

(Artetxe et al., ACL'17)

- 25 word pairs ✓
- Numeral list ✓
- Random dict. ✗



Cross-lingual embedding mappings

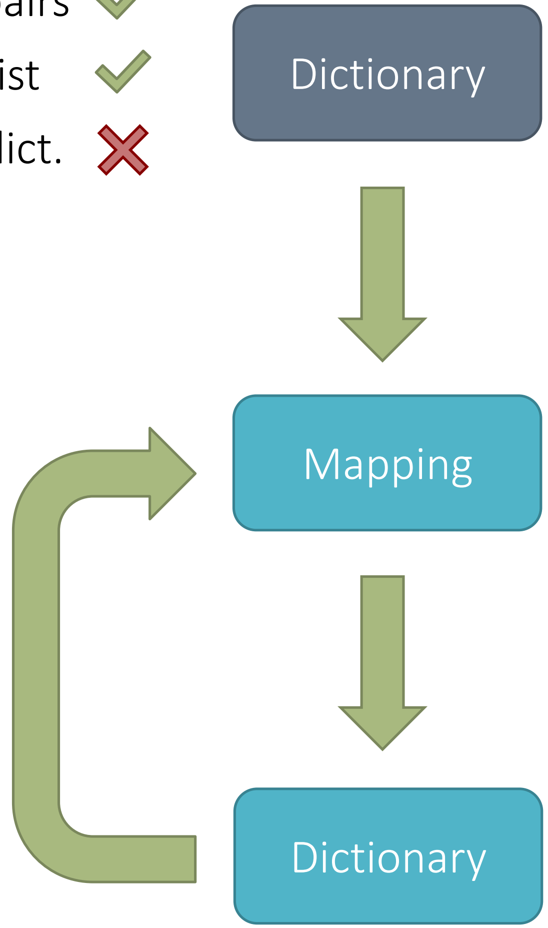


$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2$$

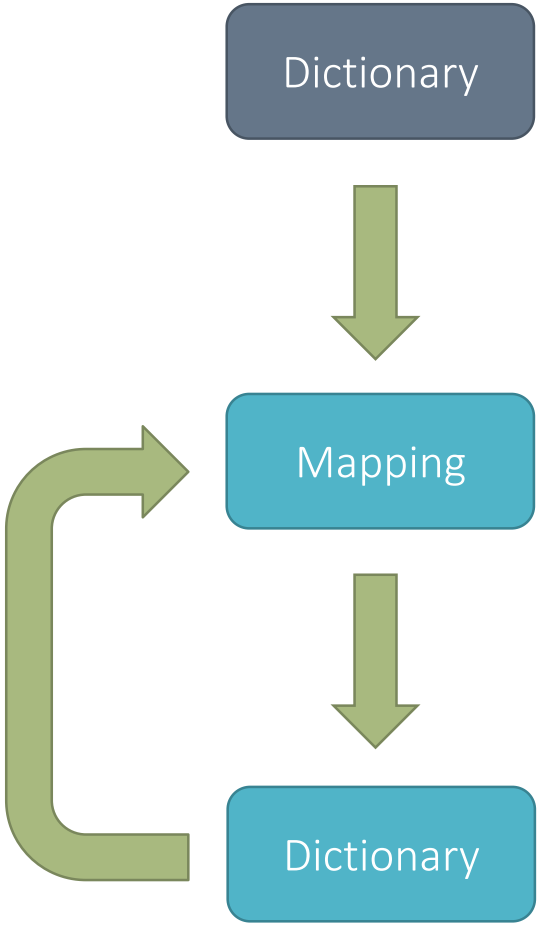
Self-learning

(Artetxe et al., ACL'17)

- 25 word pairs
- Numeral list
- Random dict.



Cross-lingual embedding mappings



<u>Intra-lingual similarity distribution</u>

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*}W - Z_{j*} \right\|^2$$

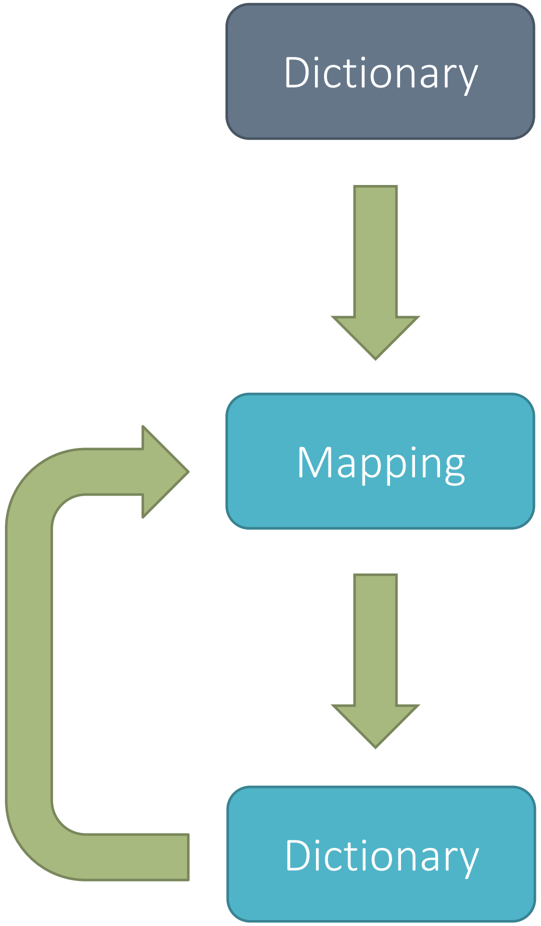
Cross-lingual embedding mappings

English

<u>Intra-lingual similarity distribution</u>

(Artetxe et al., ACL’18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



Cross-lingual embedding mappings

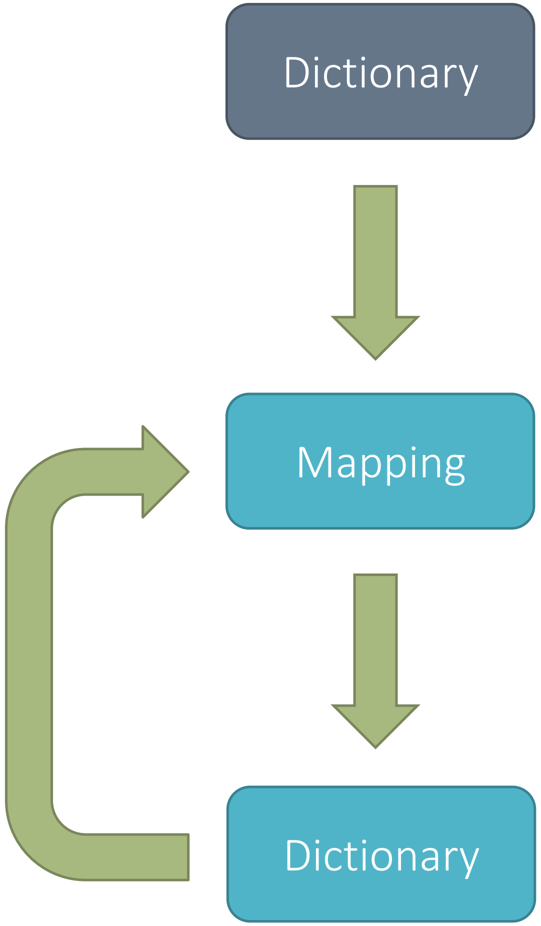
English

two

Intra-lingual similarity distribution

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



Cross-lingual embedding mappings

English

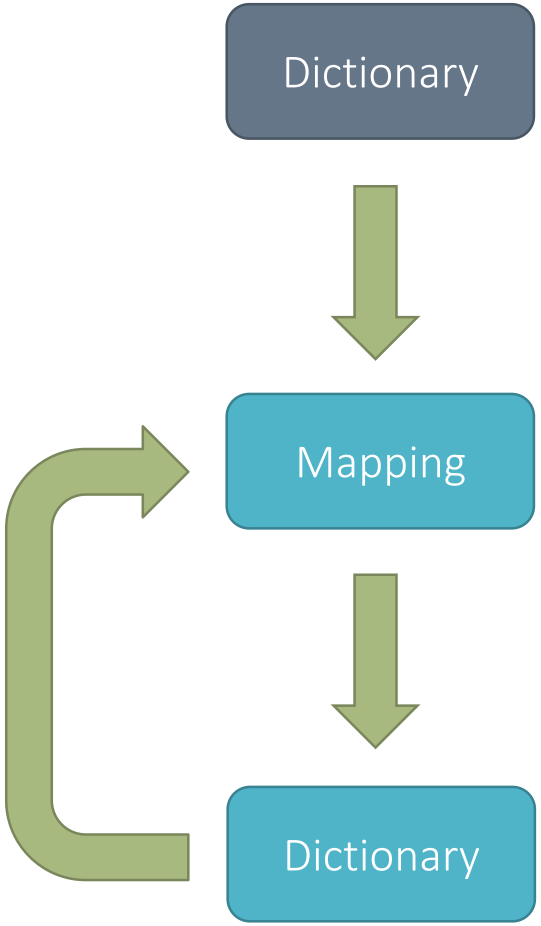
```
for x in vocab:
    sim("two", x)
```

two

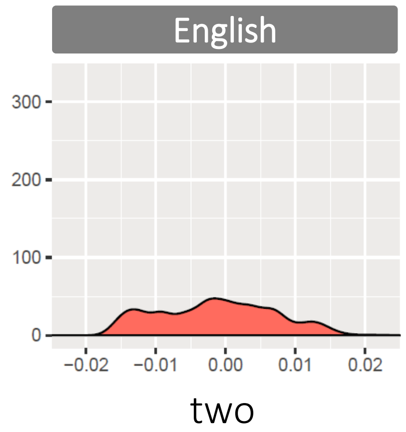
Intra-lingual similarity distribution

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



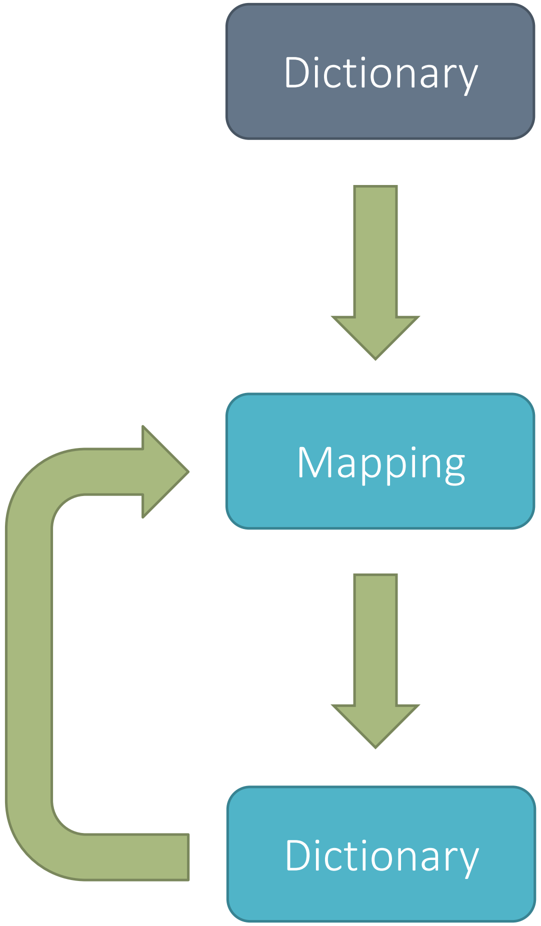
Cross-lingual embedding mappings



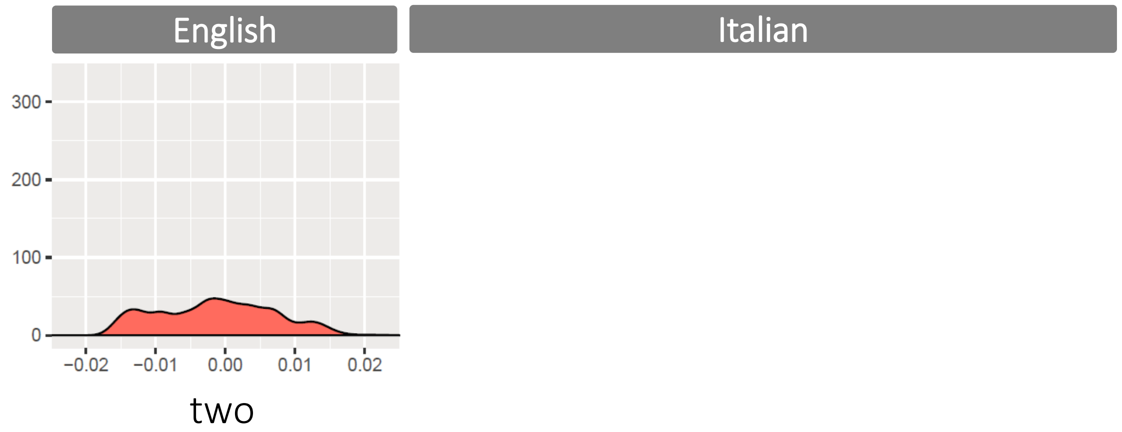
Intra-lingual similarity distribution

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*}W - Z_{j*} \right\|^2$$



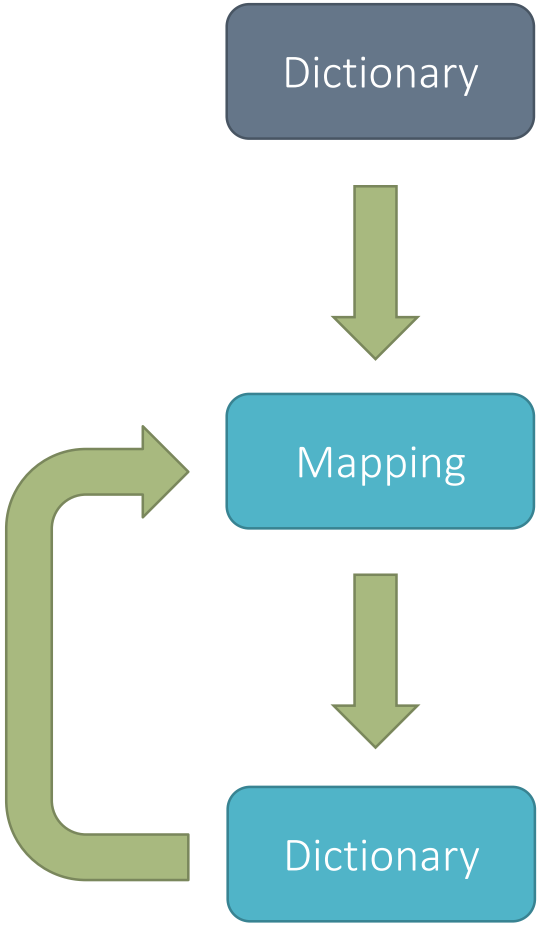
Cross-lingual embedding mappings



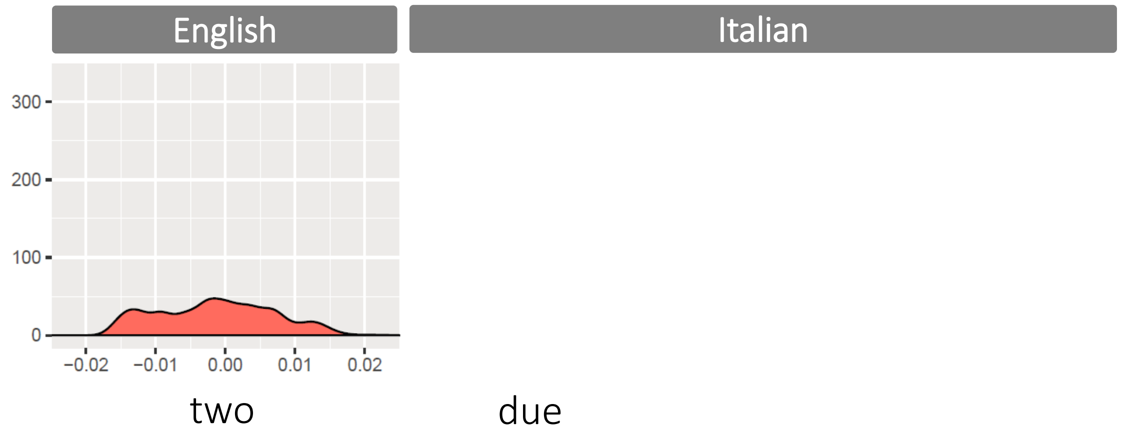
Intra-lingual similarity distribution

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



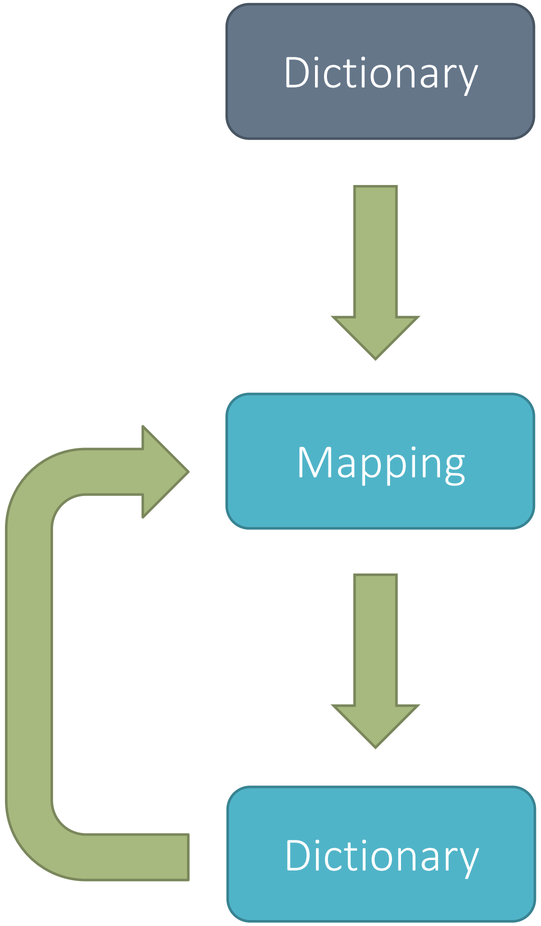
Cross-lingual embedding mappings



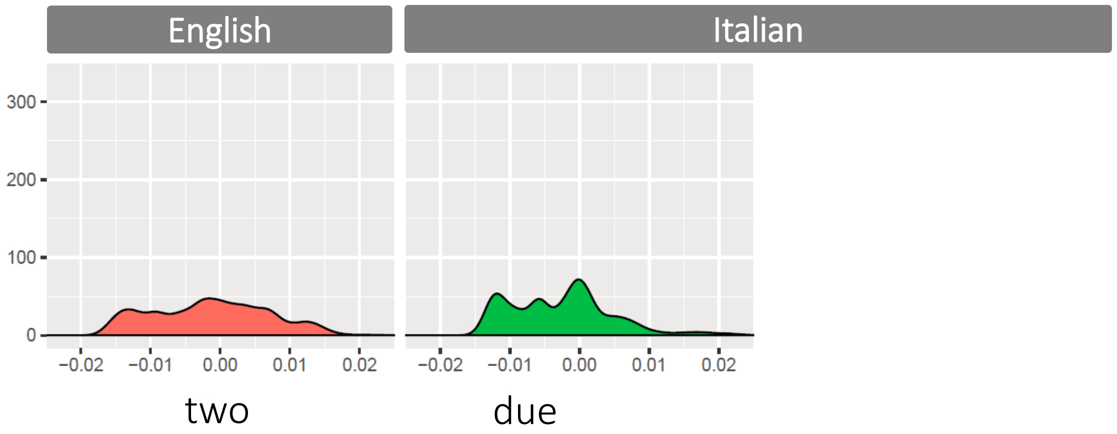
<u>Intra-lingual similarity distribution</u>

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



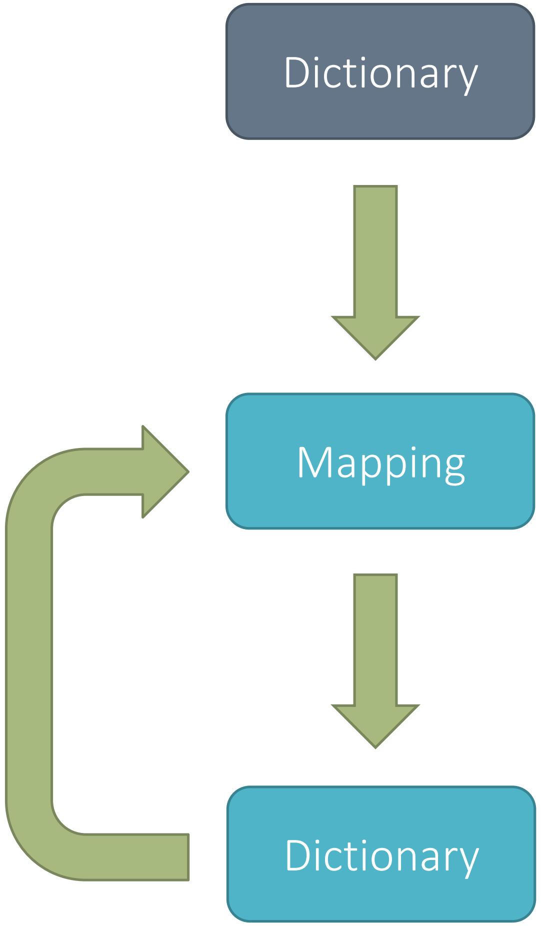
Cross-lingual embedding mappings



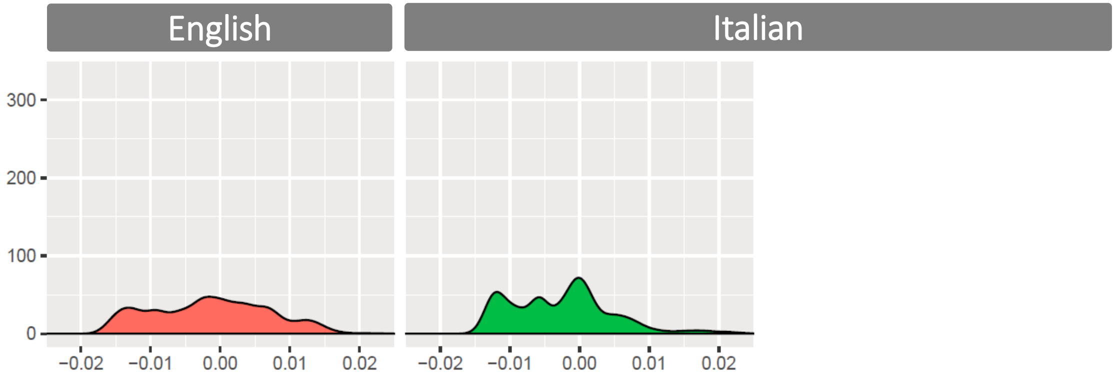
Intra-lingual similarity distribution

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



Cross-lingual embedding mappings

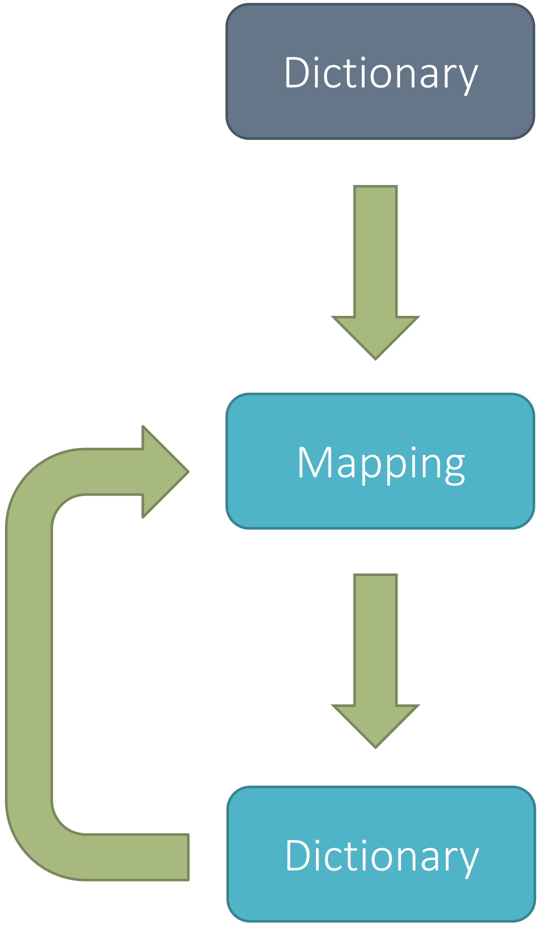


two due cane

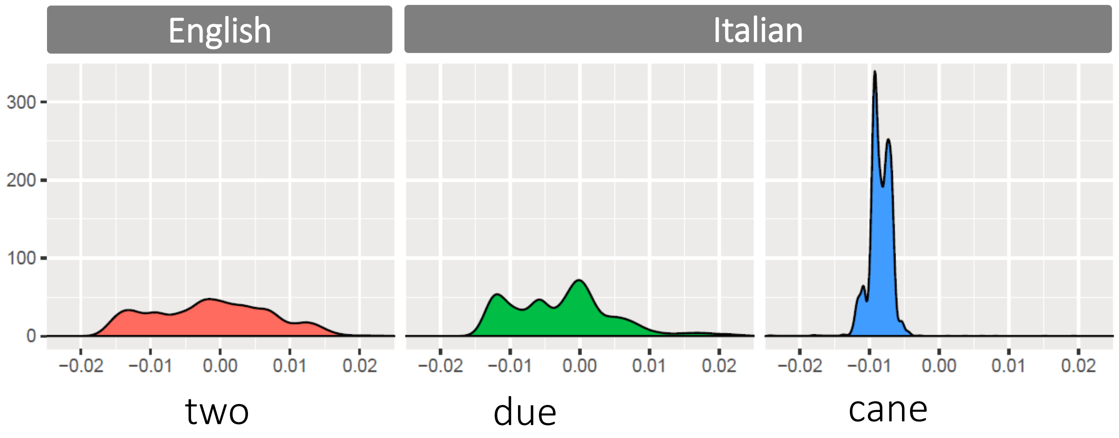
<u>Intra-lingual similarity distribution</u>

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



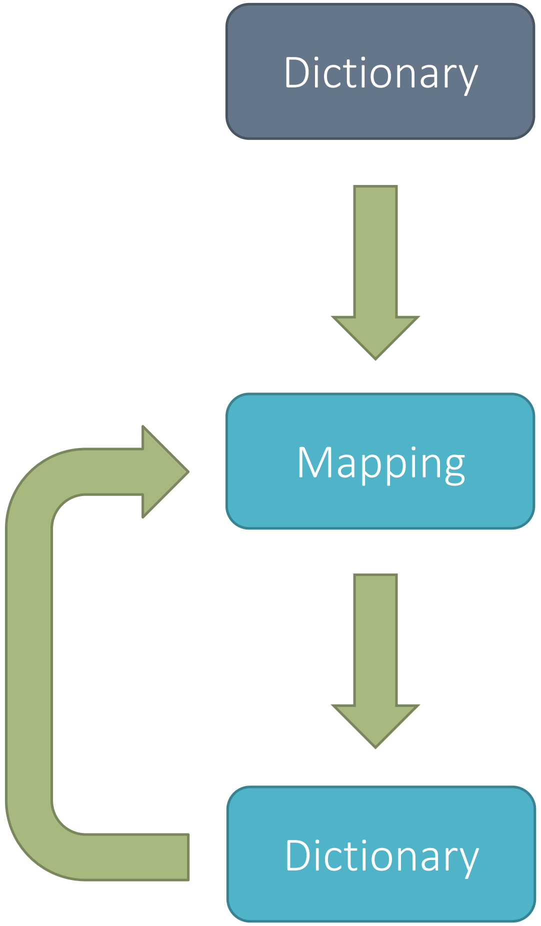
Cross-lingual embedding mappings



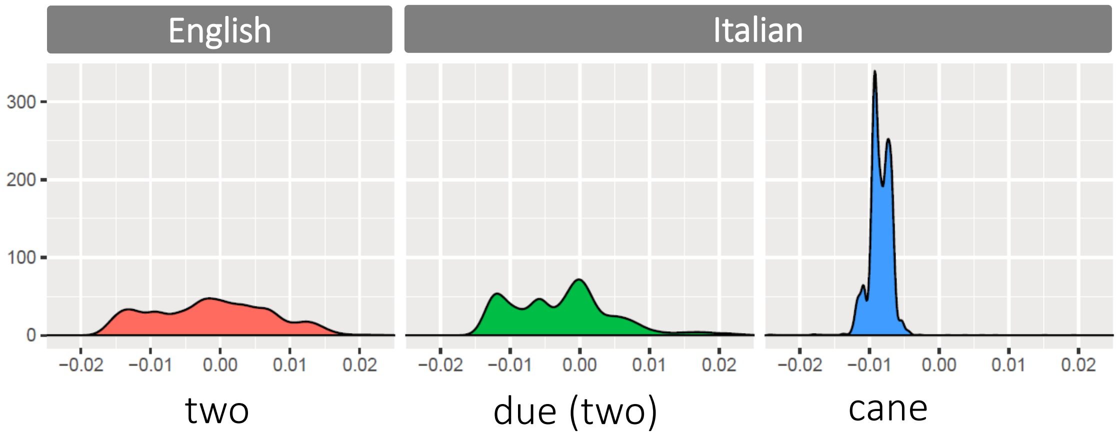
<u>Intra-lingual similarity distribution</u>

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



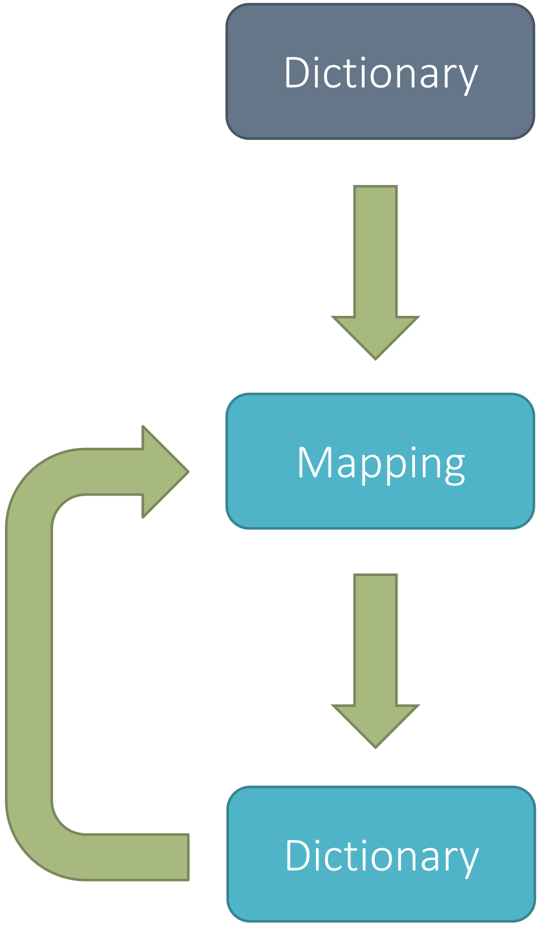
Cross-lingual embedding mappings



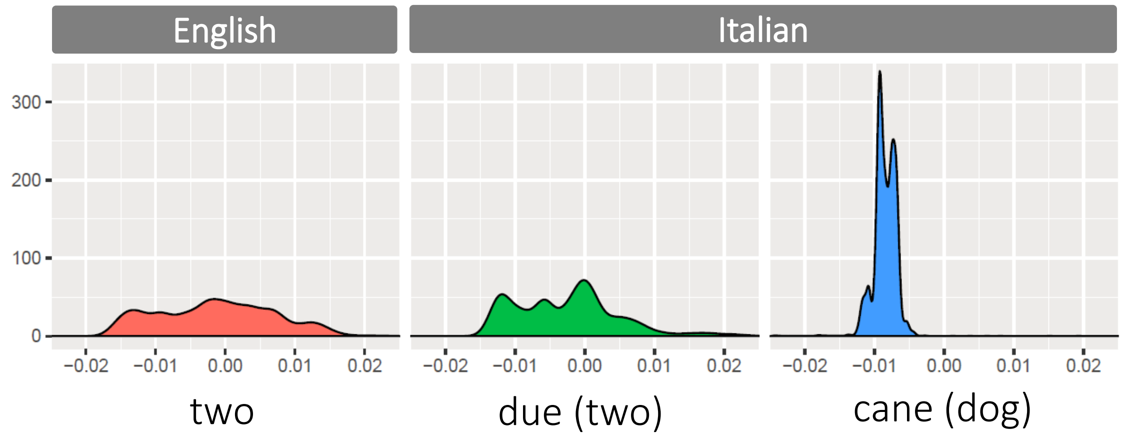
<u>**Intra-lingual similarity distribution**</u>

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \|X_{i*}W - Z_{j*}\|^2$$



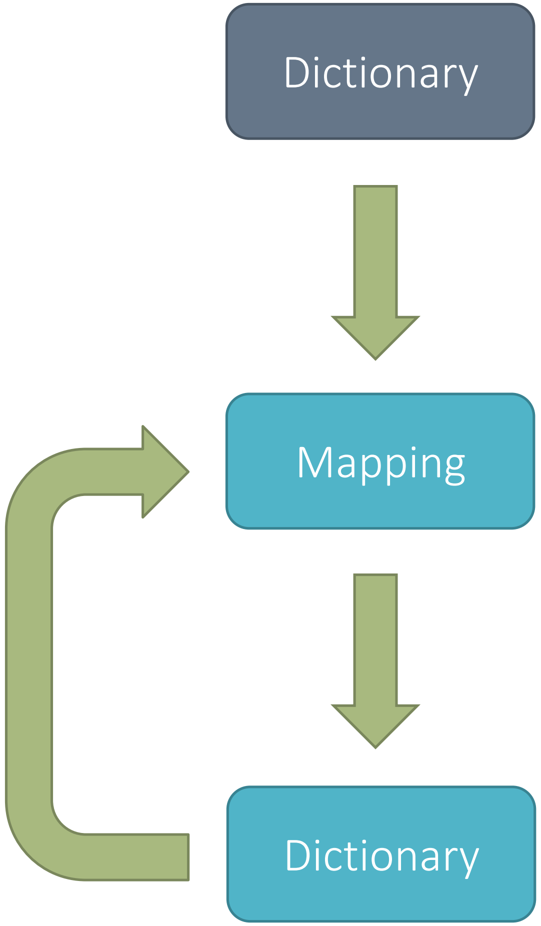
Cross-lingual embedding mappings



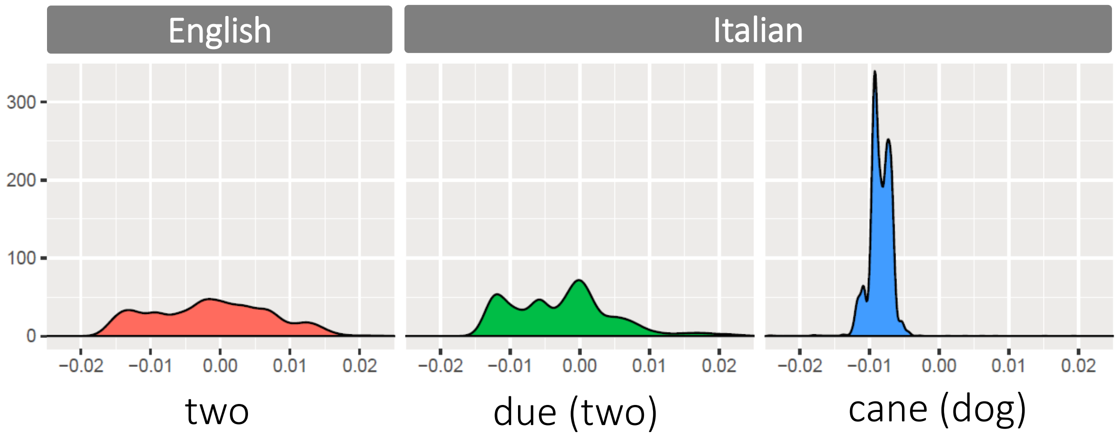
<u>**Intra-lingual similarity distribution**</u>

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



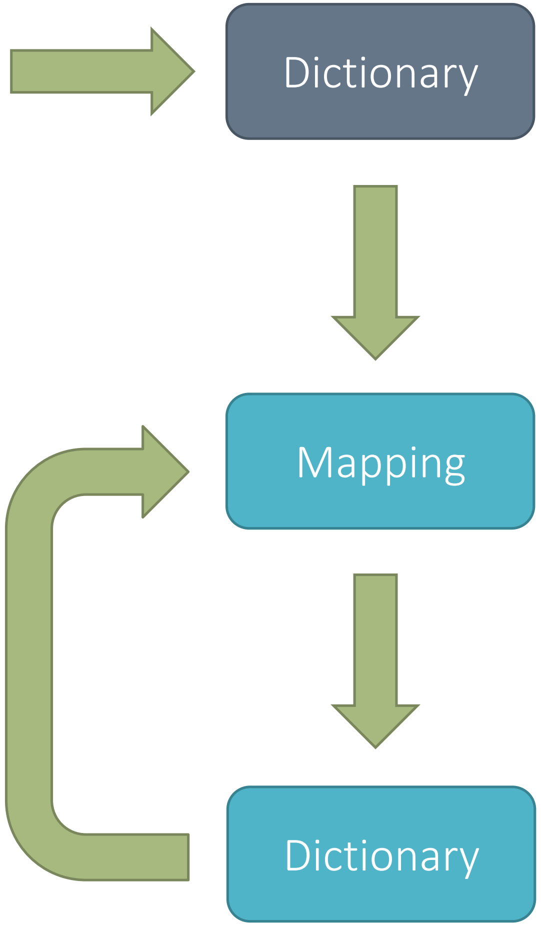
Cross-lingual embedding mappings



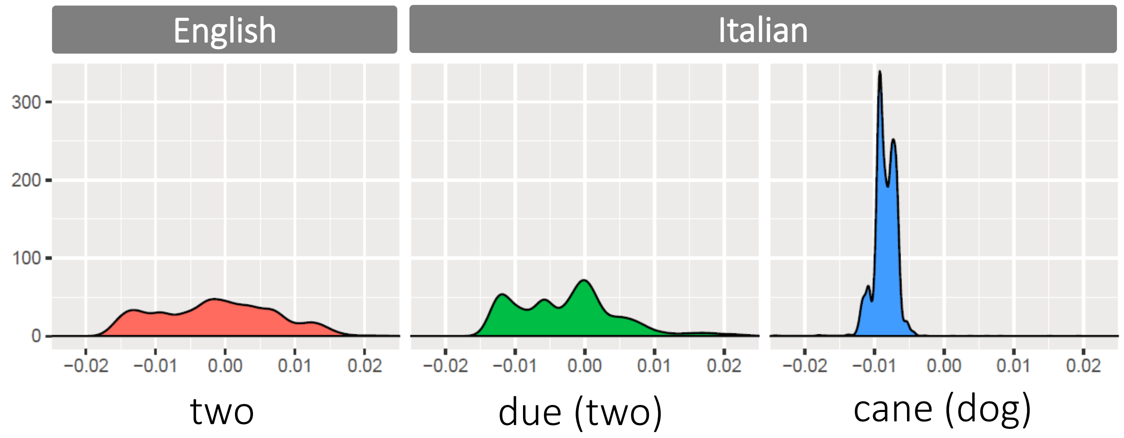
<u>Intra-lingual similarity distribution</u>

(Artetxe et al., ACL'18)

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*} W - Z_{j*} \right\|^2$$



Cross-lingual embedding mappings

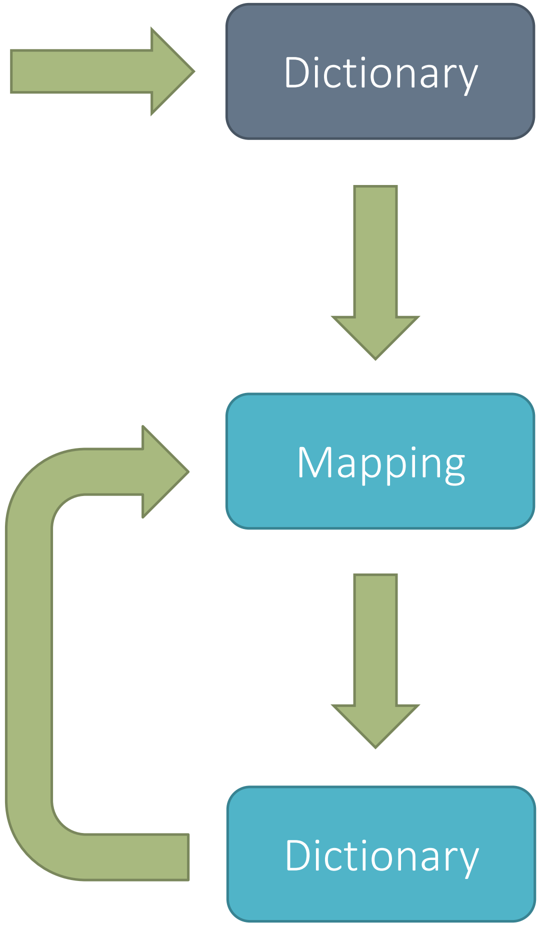


<u>Intra-lingual similarity distribution</u>

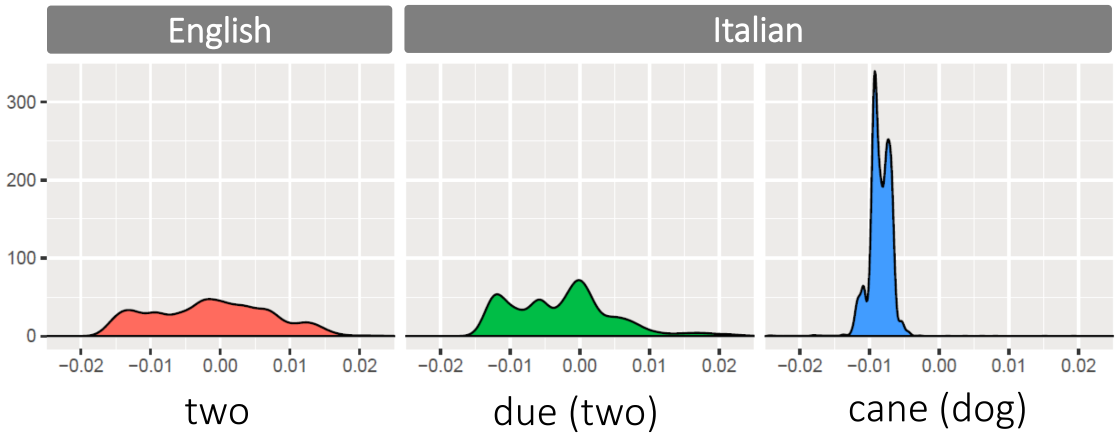
(Artetxe et al., ACL'18)

$$X' = \text{sorted}\left(\sqrt{XX^T}\right)$$

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*}W - Z_{j*} \right\|^2$$



Cross-lingual embedding mappings

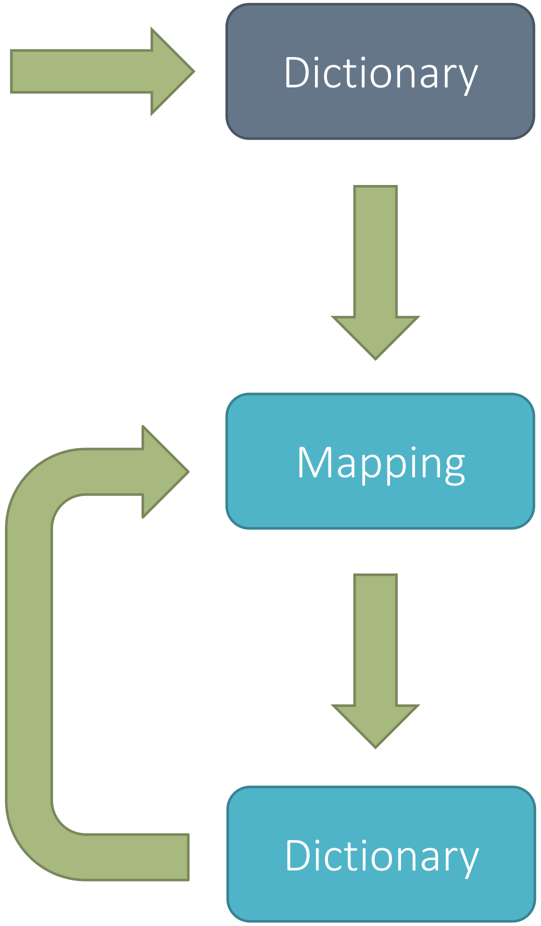


<u>Intra-lingual similarity distribution</u>

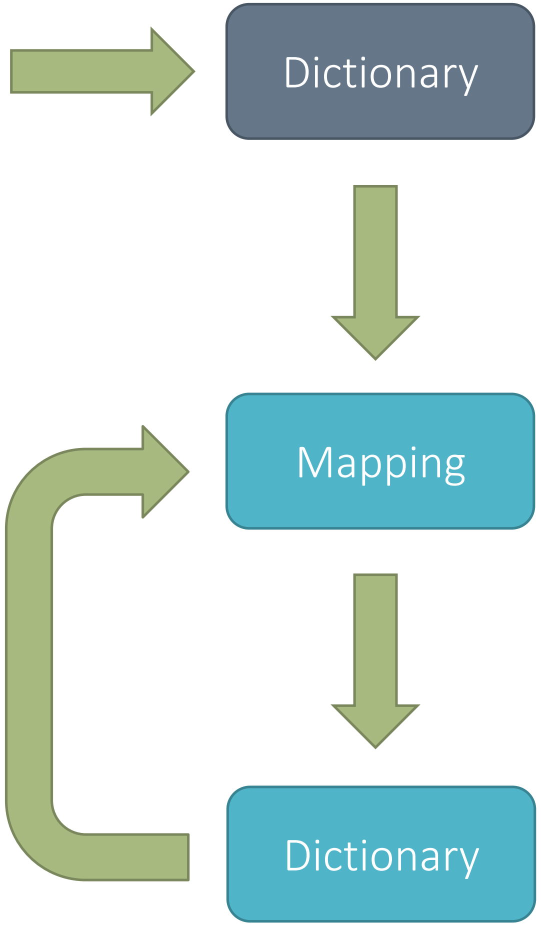
(Artetxe et al., ACL'18)

$$X' = \text{sorted}\left(\sqrt{XX^T}\right) \quad Z' = \text{sorted}\left(\sqrt{ZZ^T}\right)$$

$$W^* = \underset{W \in O(n)}{\arg\min} \sum_i \min_j \left\| X_{i*}W - Z_{j*} \right\|^2$$



Cross-lingual embedding mappings

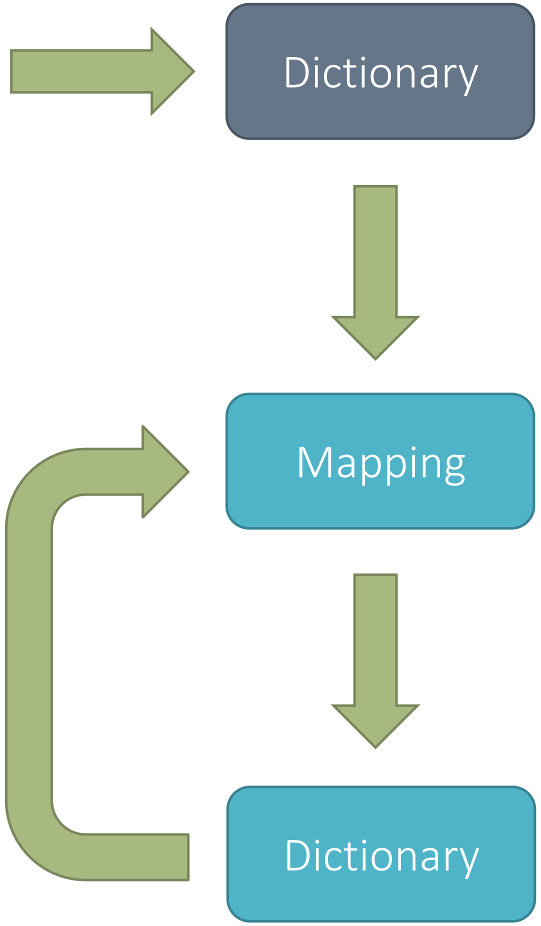


<u>**Adversarial learning**</u>

(Zhang et al, EMNLP'17;
Conneau et al., ICLR'18)

Cross-lingual embedding mappings

Generator



Adversarial learning

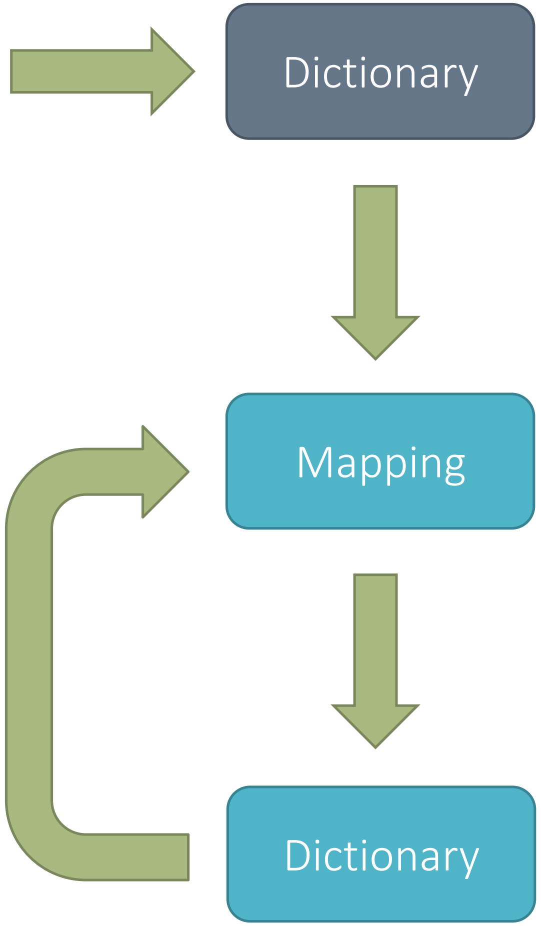
(Zhang et al, EMNLP'17;
Conneau et al., ICLR'18)

*Inspired by Li Yin's blogpost "Generative Adversarial Network"

Cross-lingual embedding mappings

Generator

Discriminator



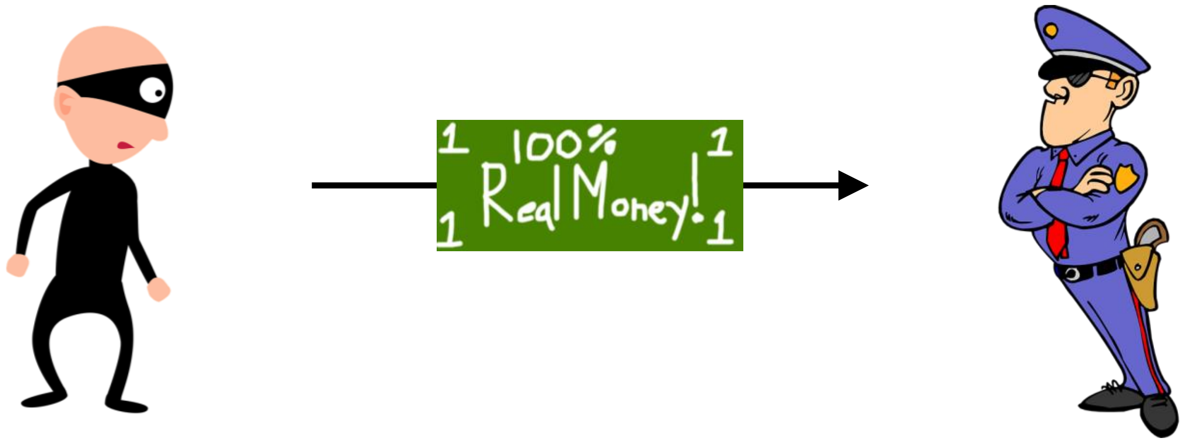
Adversarial learning

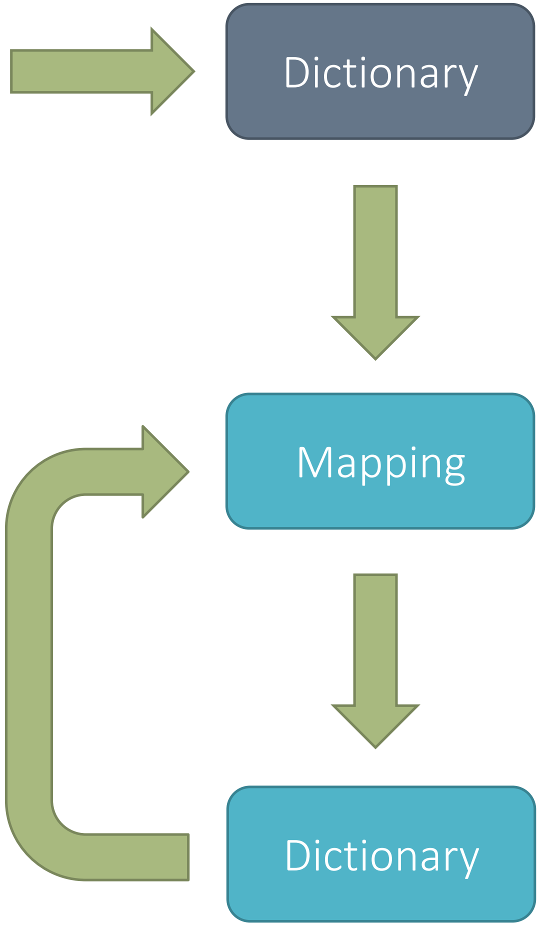
(Zhang et al, EMNLP’17;
Conneau et al., ICLR’18)

*Inspired by Li Yin’s blogpost “Generative Adversarial Network”

Cross-lingual embedding mappings

Generator **Discriminator**





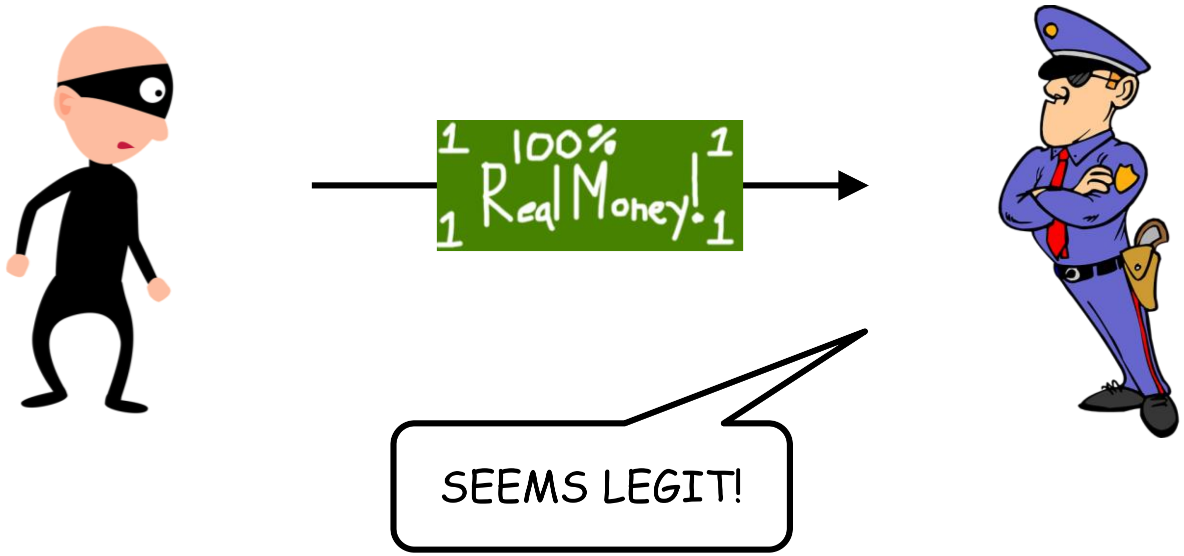
Adversarial learning

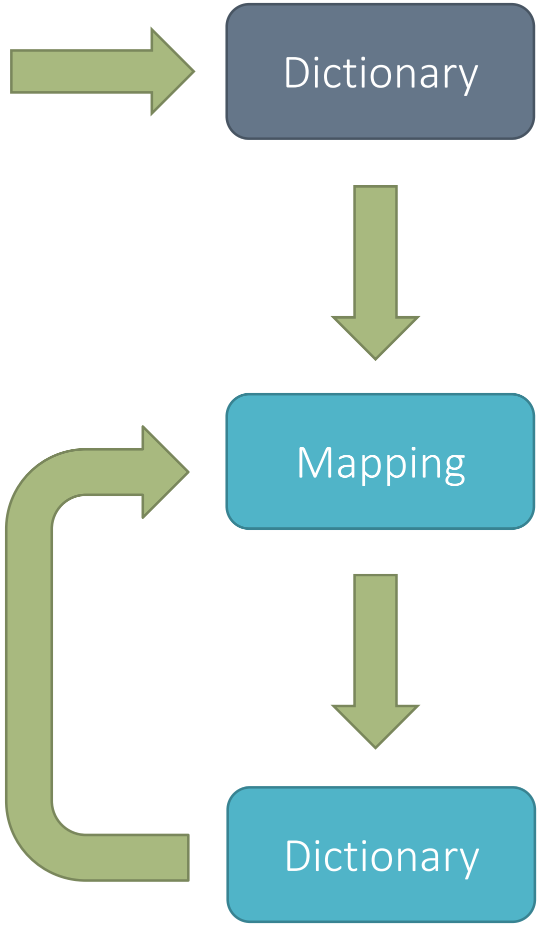
(Zhang et al, EMNLP’17;
Conneau et al., ICLR’18)

*Inspired by Li Yin’s blogpost “Generative Adversarial Network”

Cross-lingual embedding mappings

Generator **Discriminator**





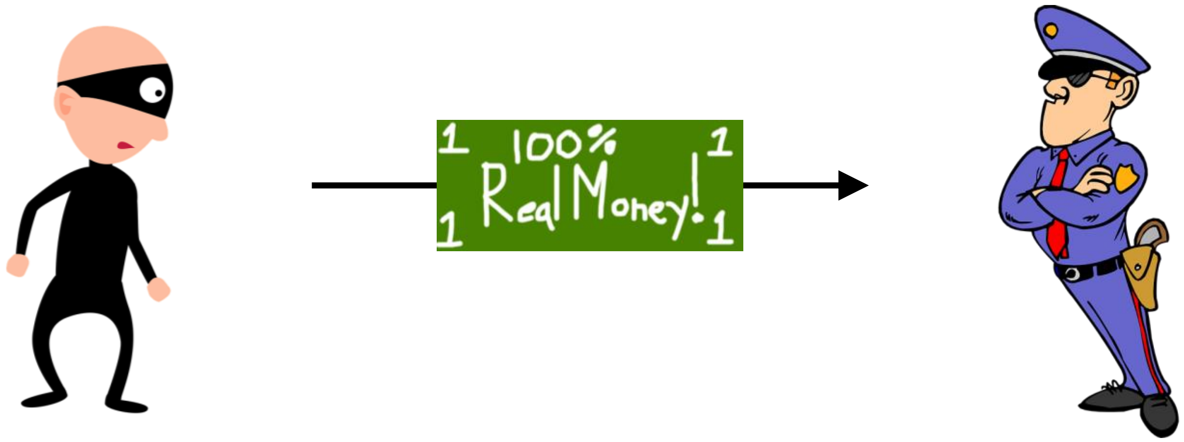
Adversarial learning

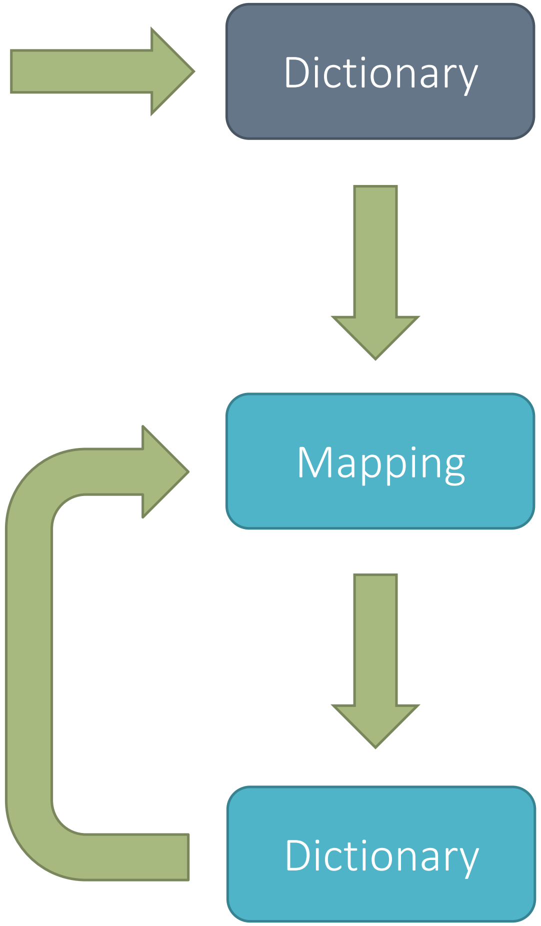
(Zhang et al, EMNLP’17;
Conneau et al., ICLR’18)

*Inspired by Li Yin’s blogpost “Generative Adversarial Network”

Cross-lingual embedding mappings

Generator **Discriminator**





Adversarial learning

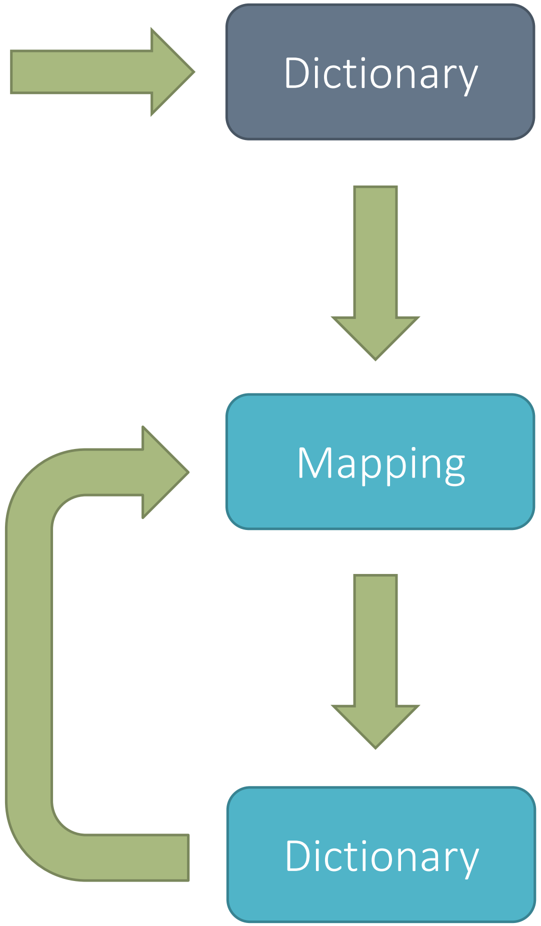
(Zhang et al, EMNLP’17;
Conneau et al., ICLR’18)

*Inspired by Li Yin’s blogpost “Generative Adversarial Network”

Cross-lingual embedding mappings

Generator

Discriminator



Adversarial learning

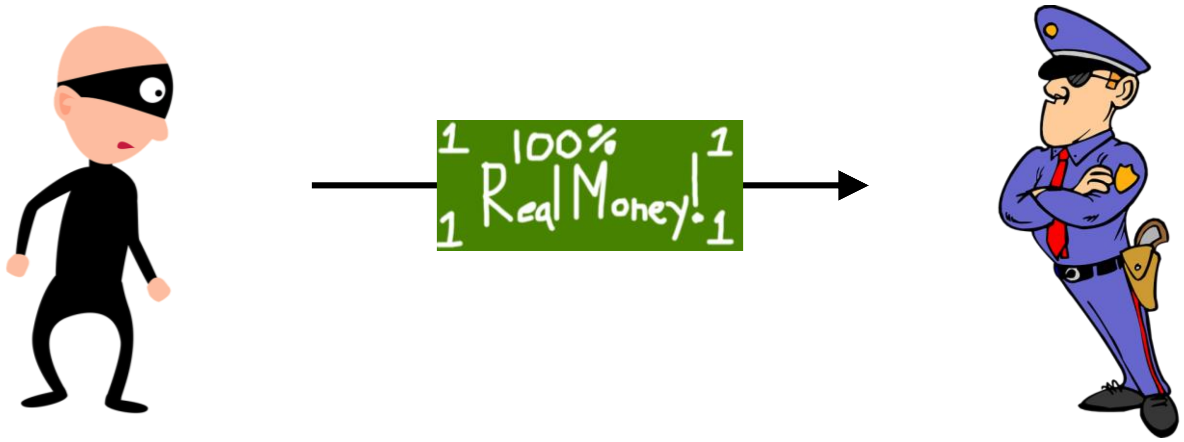
(Zhang et al, EMNLP’17;
Conneau et al., ICLR’18)

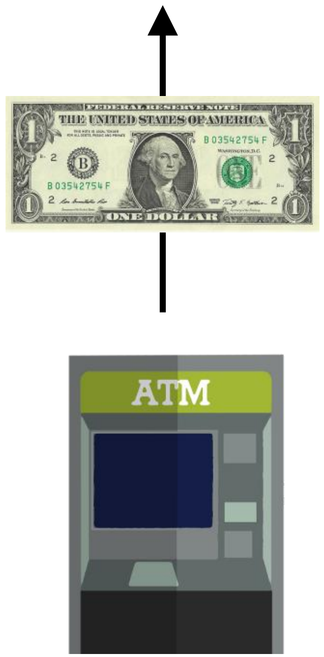
*Inspired by Li Yin’s blogpost “Generative Adversarial Network”

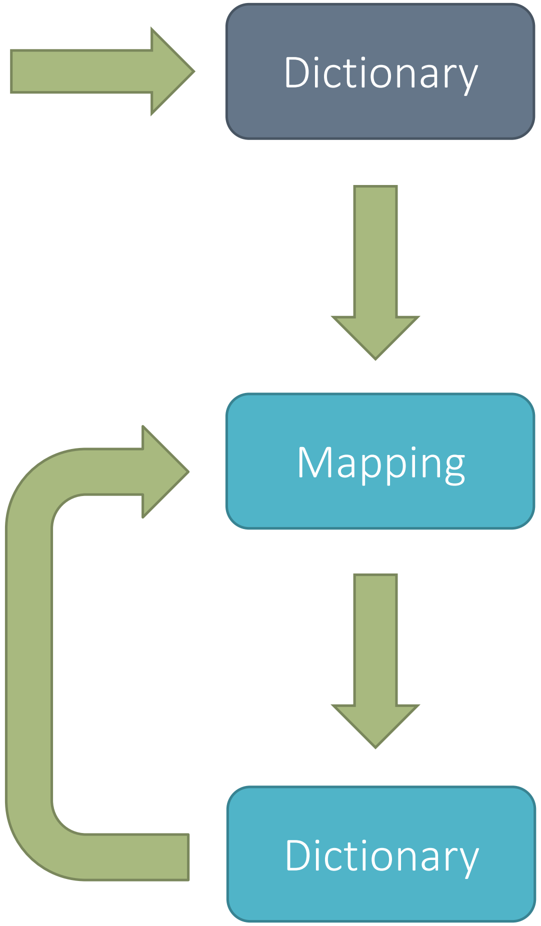
Cross-lingual embedding mappings

Generator

Discriminator







Adversarial learning

(Zhang et al, EMNLP'17;
Conneau et al., ICLR'18)

*Inspired by Li Yin's blogpost "Generative Adversarial Network"

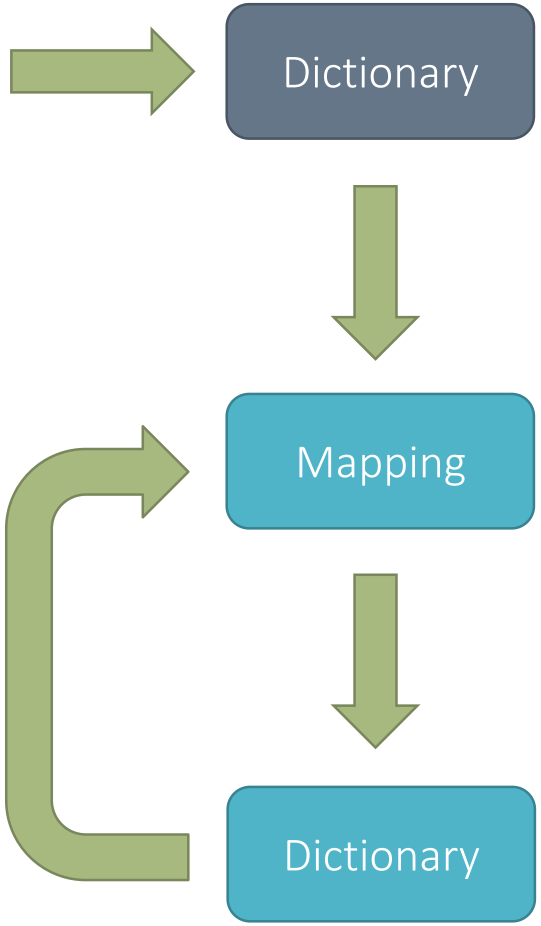
Cross-lingual embedding mappings

Generator

Discriminator

Adversarial learning

(Zhang et al, EMNLP'17;
Conneau et al., ICLR'18)

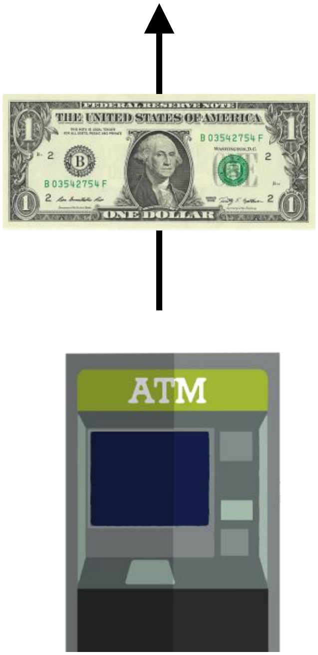


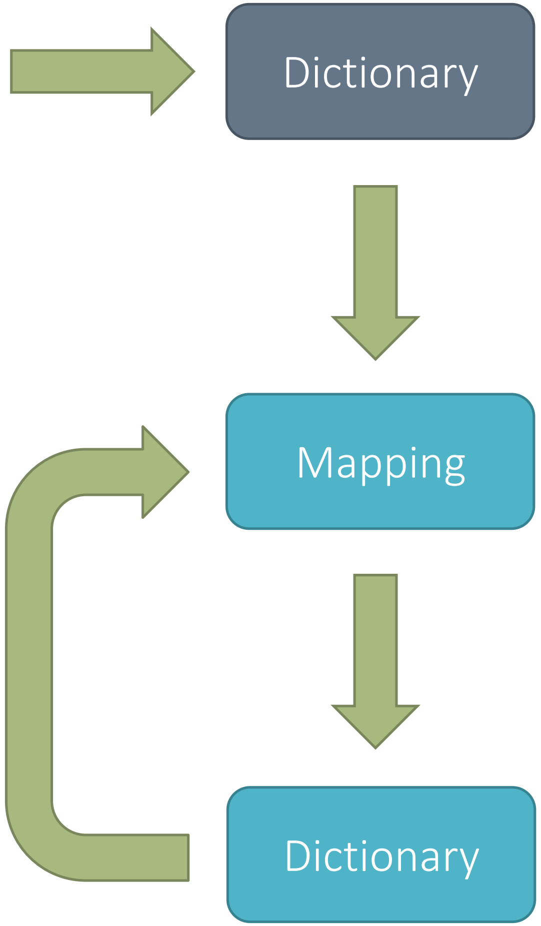
*Inspired by Li Yin's blogpost "Generative Adversarial Network"

Cross-lingual embedding mappings

Generator

Discriminator





Adversarial learning

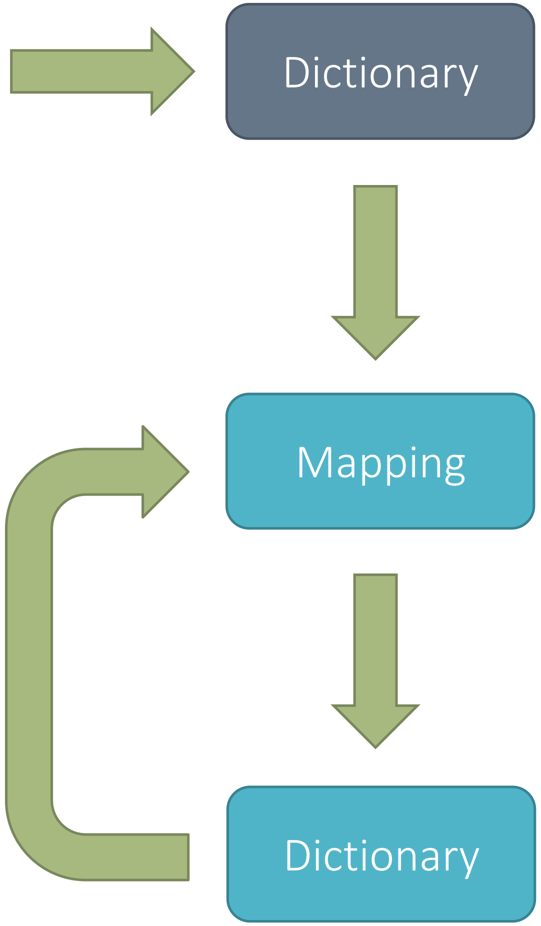
(Zhang et al, EMNLP’17;
Conneau et al., ICLR’18)

*Inspired by Li Yin’s blogpost “Generative Adversarial Network”

Cross-lingual embedding mappings

Generator

Discriminator



Adversarial learning

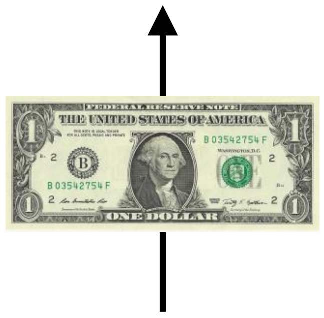
(Zhang et al, EMNLP'17;
Conneau et al., ICLR'18)

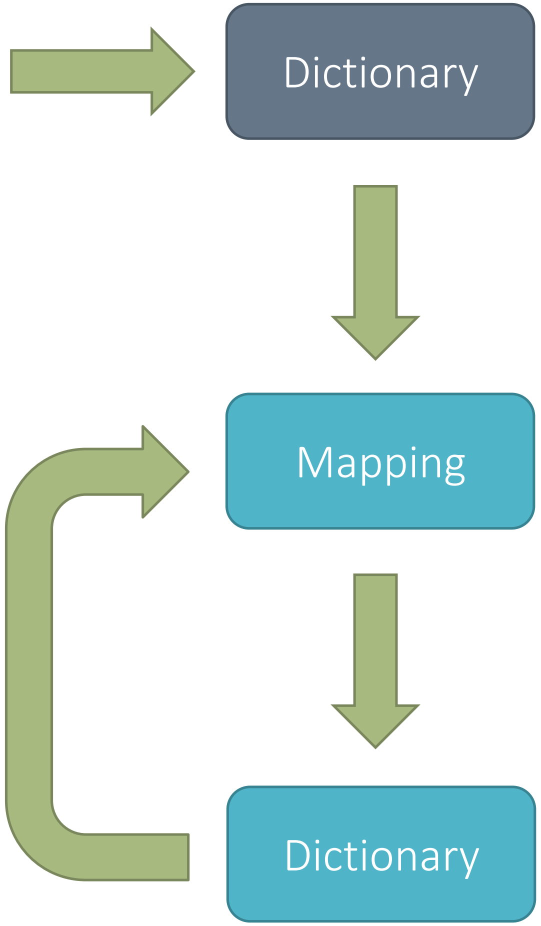
*Inspired by Li Yin's blogpost "Generative Adversarial Network"

Cross-lingual embedding mappings

Generator

Discriminator





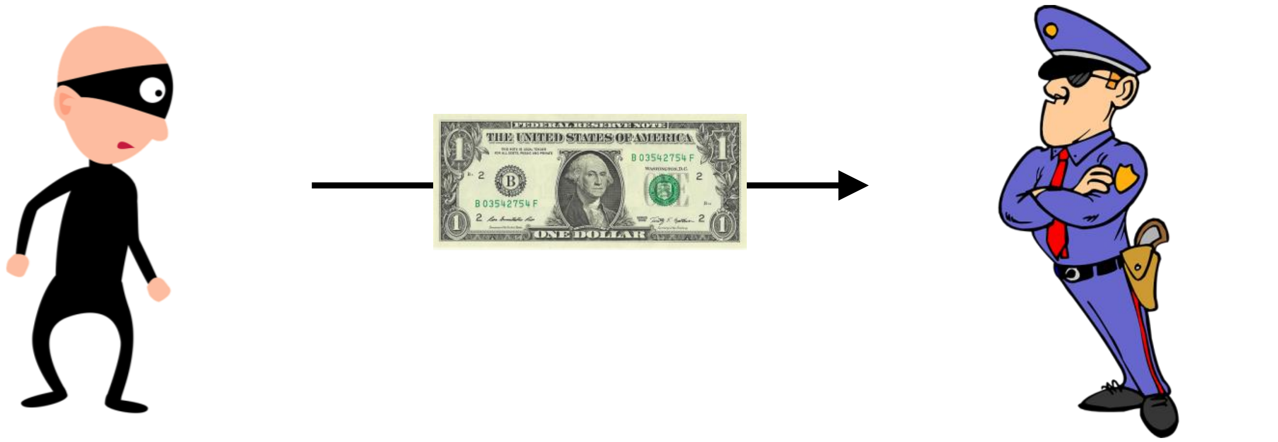
Adversarial learning

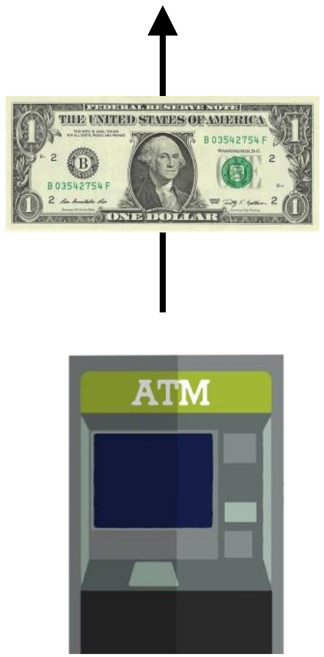
(Zhang et al, EMNLP'17;
Conneau et al., ICLR'18)

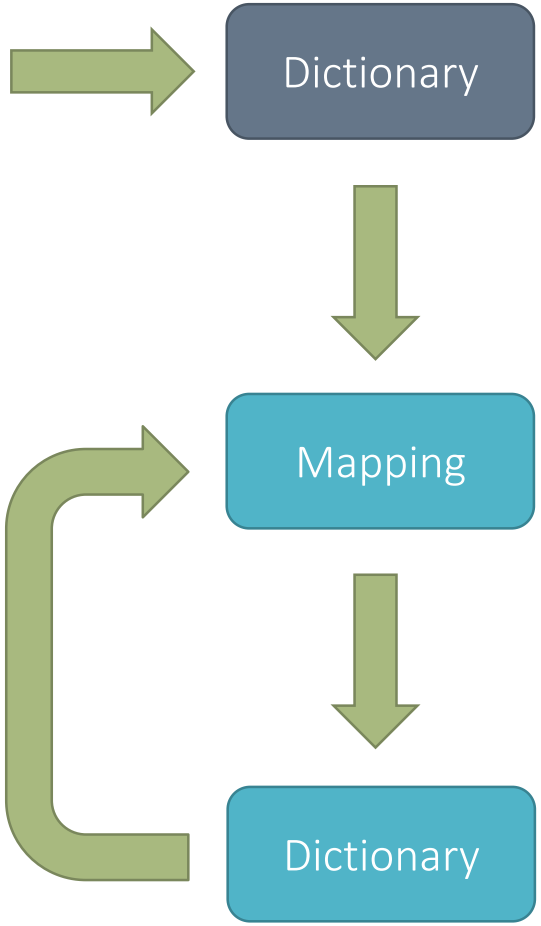
*Inspired by Li Yin's blogpost "Generative Adversarial Network"

Cross-lingual embedding mappings

Generator **Discriminator**







Adversarial learning

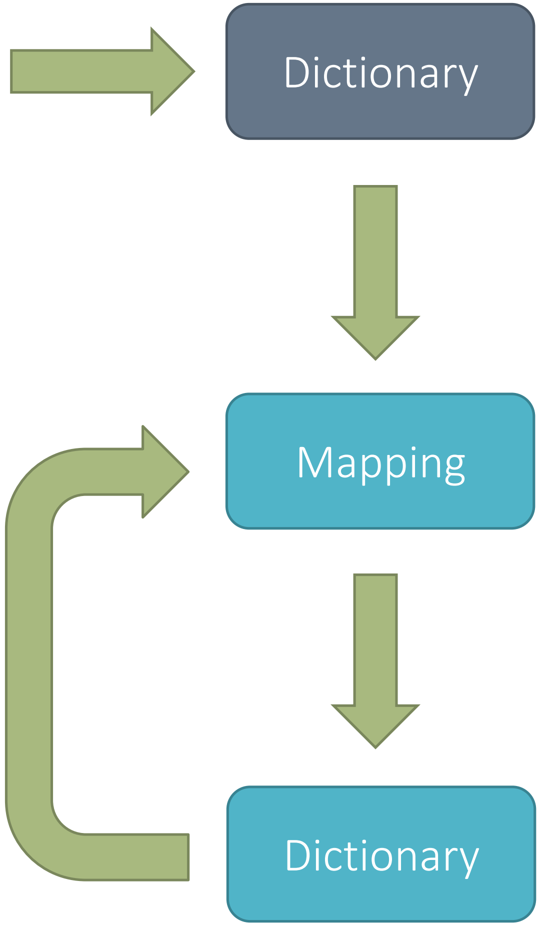
(Zhang et al, EMNLP'17;
Conneau et al., ICLR'18)

*Inspired by Li Yin's blogpost "Generative Adversarial Network"

Cross-lingual embedding mappings

Generator

Discriminator



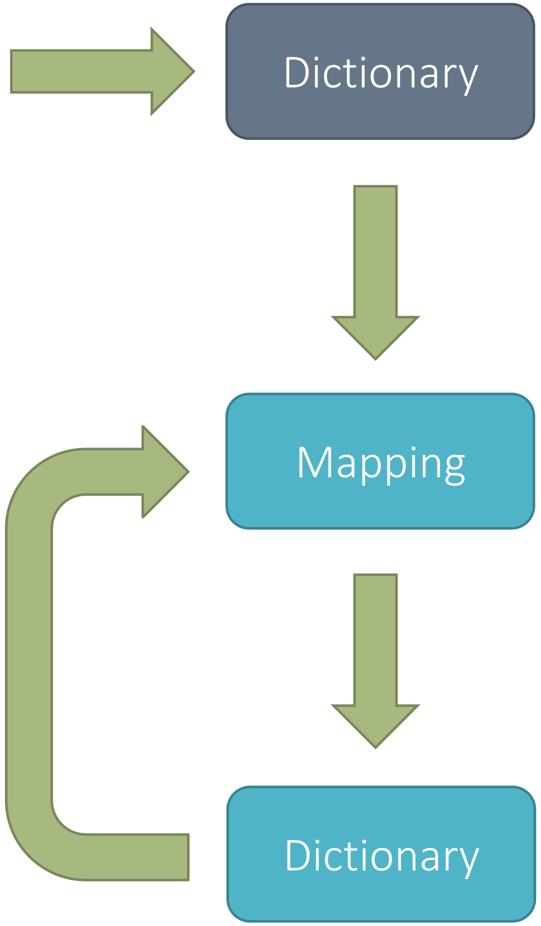
Adversarial learning

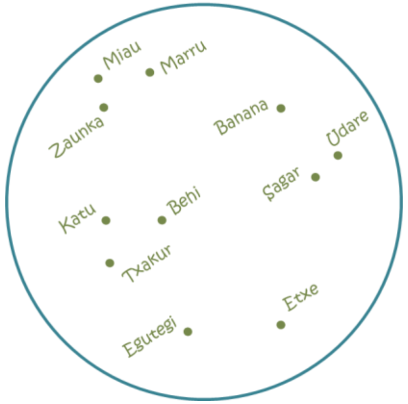
(Zhang et al, EMNLP’17;
Conneau et al., ICLR’18)

Cross-lingual embedding mappings

Generator

Discriminator





L1 embeddings

Adversarial learning

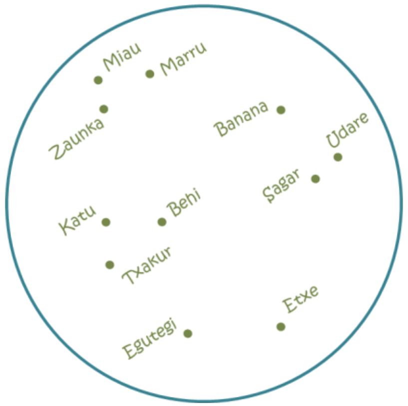
(Zhang et al, EMNLP'17;
Conneau et al., ICLR'18)

Cross-lingual embedding mappings

Generator

mapped L1 embeddings

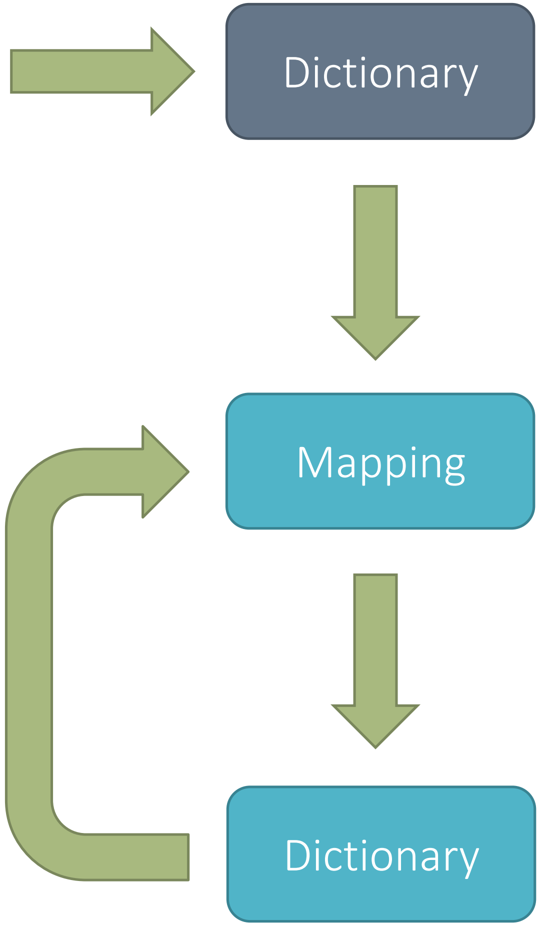
Discriminator



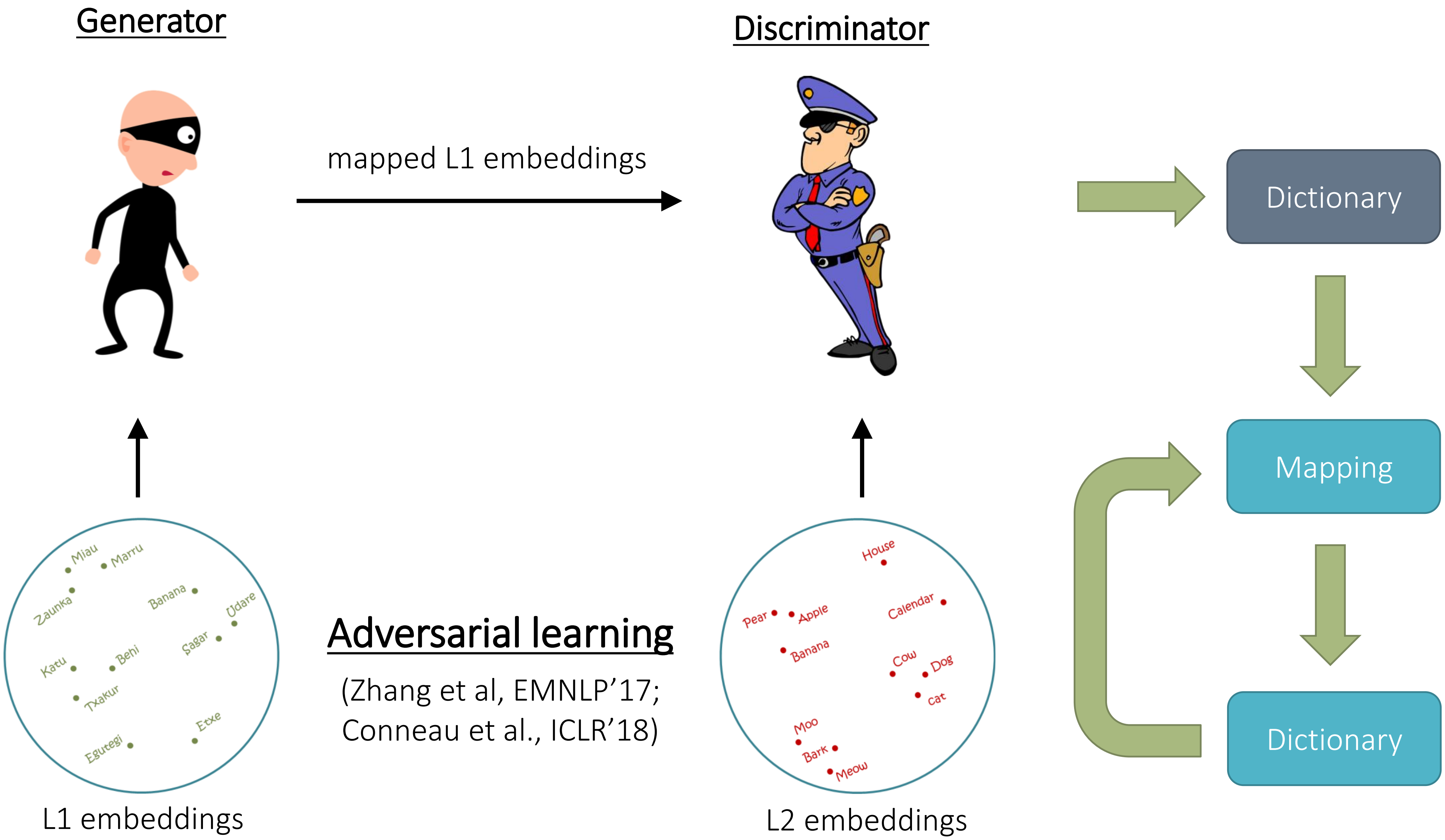
L1 embeddings

Adversarial learning

(Zhang et al, EMNLP'17;
Conneau et al., ICLR'18)



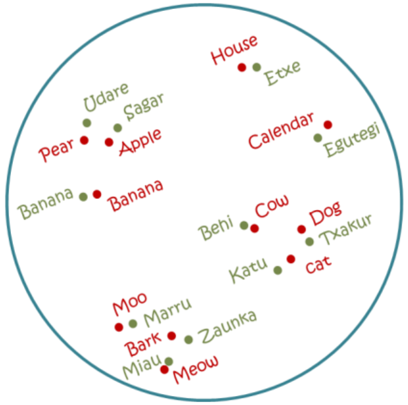
Cross-lingual embedding mappings

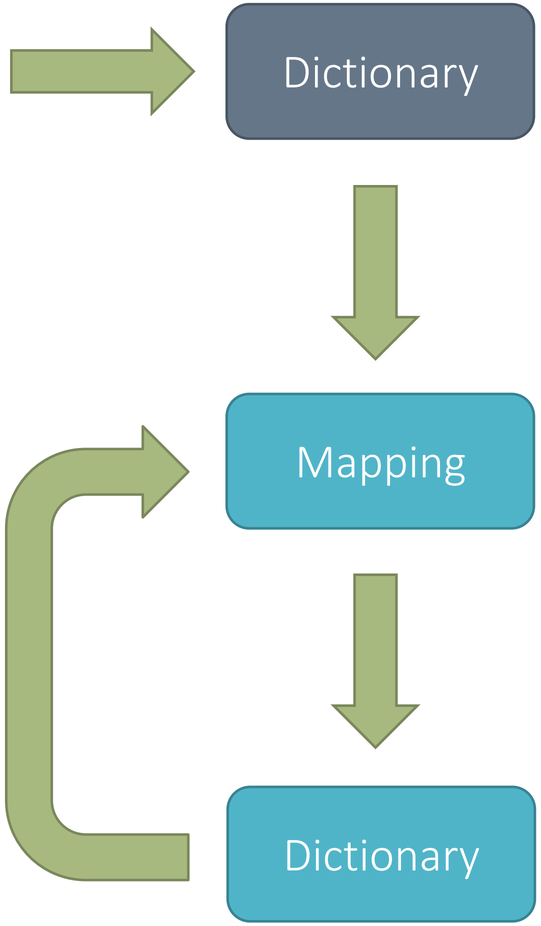


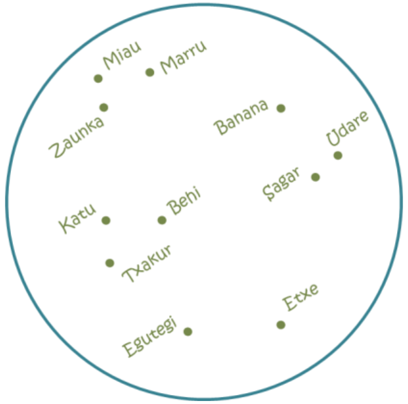
Cross-lingual embedding mappings

Generator

mapped L1 embeddings







L1 embeddings

Adversarial learning

(Zhang et al, EMNLP'17;
Conneau et al., ICLR'18)

Cross-lingual embedding mappings

Cross-lingual embedding mappings

Supervision	Method

Cross-lingual embedding mappings

Supervision	Method
5k dict.	Mikolov et al. (2013)
	Faruqui and Dyer (2014)
	Shigeto et al. (2015)
	Dinu et al. (2015)
	Lazaridou et al. (2015)
	Xing et al. (2015)
	Zhang et al. (2016)
	Artetxe et al. (2016)
	Smith et al. (2017)
	Artetxe et al. (2018a)

Cross-lingual embedding mappings

Supervision	Method
5k dict.	Mikolov et al. (2013)
	Faruqui and Dyer (2014)
	Shigeto et al. (2015)
	Dinu et al. (2015)
	Lazaridou et al. (2015)
	Xing et al. (2015)
	Zhang et al. (2016)
	Artetxe et al. (2016)
	Smith et al. (2017)
	Artetxe et al. (2018a)
25 dict.	**Artetxe et al. (2017)**

Cross-lingual embedding mappings

Supervision	Method
5k dict.	Mikolov et al. (2013)
	Faruqui and Dyer (2014)
	Shigeto et al. (2015)
	Dinu et al. (2015)
	Lazaridou et al. (2015)
	Xing et al. (2015)
	Zhang et al. (2016)
	Artetxe et al. (2016)
	Smith et al. (2017)
	Artetxe et al. (2018a)
25 dict.	**Artetxe et al. (2017)**
Init. heurist.	Smith et al. (2017), cognates
	Artetxe et al. (2017), num.

Cross-lingual embedding mappings

Supervision	Method
5k dict.	Mikolov et al. (2013)
	Faruqui and Dyer (2014)
	Shigeto et al. (2015)
	Dinu et al. (2015)
	Lazaridou et al. (2015)
	Xing et al. (2015)
	Zhang et al. (2016)
	Artetxe et al. (2016)
	Smith et al. (2017)
	Artetxe et al. (2018a)
25 dict.	**Artetxe et al. (2017)**
Init. heurist.	Smith et al. (2017), cognates
	Artetxe et al. (2017), num.
None	Zhang et al. (2017), $\lambda = 1$
	Zhang et al. (2017), $\lambda = 10$
	Conneau et al. (2018), code‡
	Conneau et al. (2018), paper‡
	Artetxe et al. (2018b)

Cross-lingual embedding mappings

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)				
	Faruqui and Dyer (2014)				
	Shigeto et al. (2015)				
	Dinu et al. (2015)				
	Lazaridou et al. (2015)				
	Xing et al. (2015)				
	Zhang et al. (2016)				
	Artetxe et al. (2016)				
	Smith et al. (2017)				
	Artetxe et al. (2018a)				
25 dict.	**Artetxe et al. (2017)**				
Init. heurist.	Smith et al. (2017), cognates				
	Artetxe et al. (2017), num.				
None	Zhang et al. (2017), $\lambda = 1$				
	Zhang et al. (2017), $\lambda = 10$				
	Conneau et al. (2018), code[‡]				
	Conneau et al. (2018), paper[‡]				
	Artetxe et al. (2018b)				

Cross-lingual embedding mappings

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)	34.93†	35.00†	25.91†	27.73†
	Faruqui and Dyer (2014)	38.40*	37.13*	27.60*	26.80*
	Shigeto et al. (2015)	41.53†	43.07†	31.04†	33.73†
	Dinu et al. (2015)	37.7	38.93*	29.14*	30.40*
	Lazaridou et al. (2015)	40.2	-	-	-
	Xing et al. (2015)	36.87†	41.27†	28.23†	31.20†
	Zhang et al. (2016)	36.73†	40.80†	28.16†	31.07†
	Artetxe et al. (2016)	39.27	41.87*	30.62*	31.40*
	Smith et al. (2017)	43.1	43.33†	29.42†	35.13†
	Artetxe et al. (2018a)	45.27	44.13	**32.94**	36.60
25 dict.	**Artetxe et al. (2017)**	37.27	39.60	28.16	-
Init. heurist.	Smith et al. (2017), cognates	39.9	-	-	-
	Artetxe et al. (2017), num.	39.40	40.27	26.47	-
None	Zhang et al. (2017), $\lambda = 1$	0.00*	0.00*	0.00*	0.00*
	Zhang et al. (2017), $\lambda = 10$	0.00*	0.00*	0.01*	0.01*
	Conneau et al. (2018), code‡	45.15*	46.83*	0.38*	35.38*
	Conneau et al. (2018), paper‡	45.1	0.01*	0.01*	35.44*
	Artetxe et al. (2018b)	**48.13**	**48.19**	32.63	**37.33**

Cross-lingual embedding mappings

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)	34.93†	35.00†	25.91†	27.73†
	Faruqui and Dyer (2014)	38.40*	37.13*	27.60*	26.80*
	Shigeto et al. (2015)	41.53†	43.07†	31.04†	33.73†
	Dinu et al. (2015)	37.7	38.93*	29.14*	30.40*
	Lazaridou et al. (2015)	40.2	-	-	-
	Xing et al. (2015)	36.87†	41.27†	28.23†	31.20†
	Zhang et al. (2016)	36.73†	40.80†	28.16†	31.07†
	Artetxe et al. (2016)	39.27	41.87*	30.62*	31.40*
	Smith et al. (2017)	43.1	43.33†	29.42†	35.13†
	Artetxe et al. (2018a)	45.27	44.13	**32.94**	36.60
25 dict.	**Artetxe et al. (2017)**	37.27	39.60	28.16	-
Init. heurist.	Smith et al. (2017), cognates	39.9	-	-	-
	Artetxe et al. (2017), num.	39.40	40.27	26.47	-
None	Zhang et al. (2017), $\lambda = 1$	0.00*	0.00*	0.00*	0.00*
	Zhang et al. (2017), $\lambda = 10$	0.00*	0.00*	0.01*	0.01*
	Conneau et al. (2018), code‡	45.15*	46.83*	0.38*	35.38*
	Conneau et al. (2018), paper‡	45.1	0.01*	0.01*	35.44*
	Artetxe et al. (2018b)	**48.13**	**48.19**	32.63	**37.33**

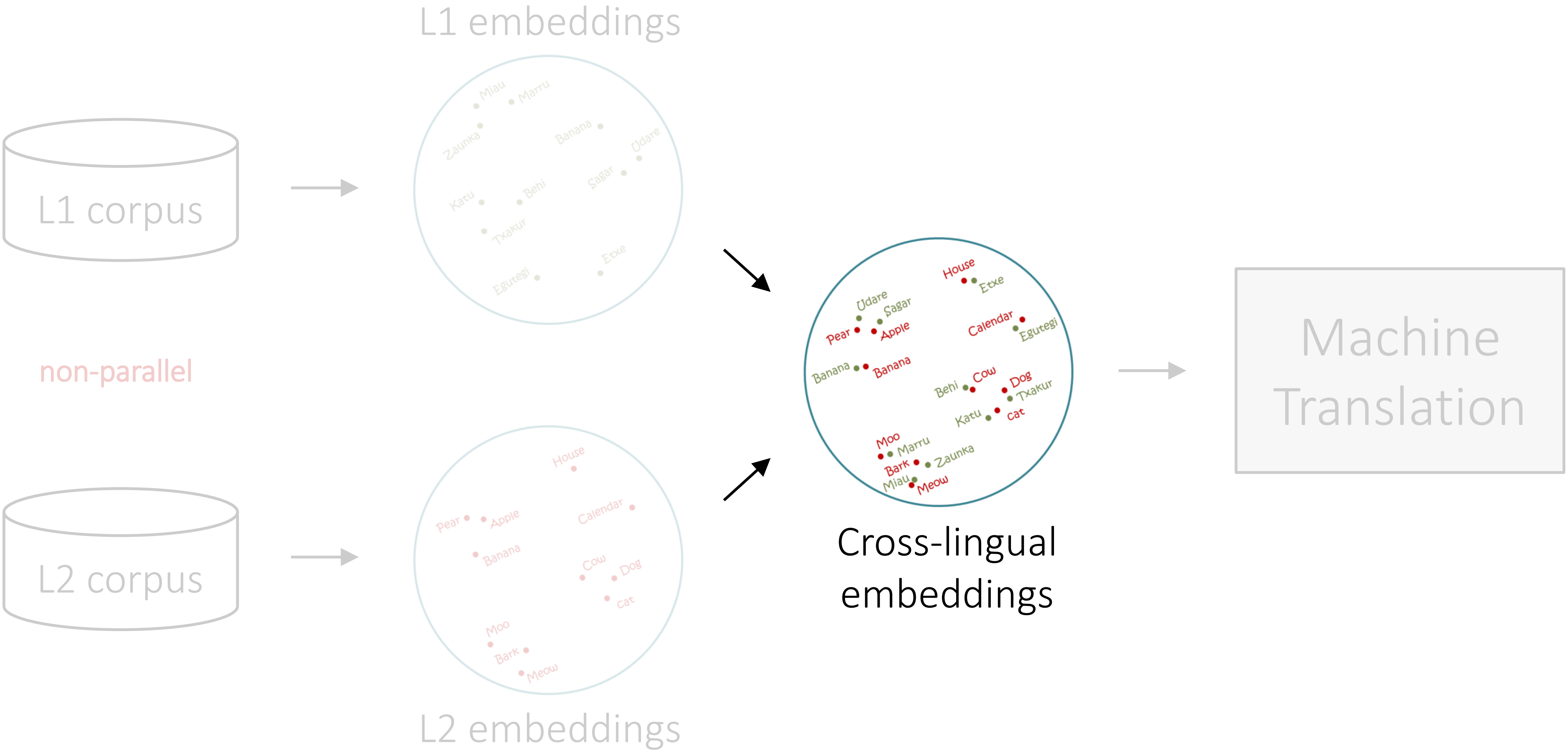
Cross-lingual embedding mappings

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)	34.93†	35.00†	25.91†	27.73†
	Faruqui and Dyer (2014)	38.40*	37.13*	27.60*	26.80*
	Shigeto et al. (2015)	41.53†	43.07†	31.04†	33.73†
	Dinu et al. (2015)	37.7	38.93*	29.14*	30.40*
	Lazaridou et al. (2015)	40.2	-	-	-
	Xing et al. (2015)	36.87†	41.27†	28.23†	31.20†
	Zhang et al. (2016)	36.73†	40.80†	28.16†	31.07†
	Artetxe et al. (2016)	39.27	41.87*	30.62*	31.40*
	Smith et al. (2017)	43.1	43.33†	29.42†	35.13†
	Artetxe et al. (2018a)	45.27	44.13	32.94	36.60
25 dict.	**Artetxe et al. (2017)**	37.27	39.60	28.16	-
Init. heurist.	Smith et al. (2017), cognates	39.9	-	-	-
	Artetxe et al. (2017), num.	39.40	40.27	26.47	-
None	Zhang et al. (2017), $\lambda = 1$	0.00*	0.00*	0.00*	0.00*
	Zhang et al. (2017), $\lambda = 10$	0.00*	0.00*	0.01*	0.01*
	Conneau et al. (2018), code‡	45.15*	46.83*	0.38*	35.38*
	Conneau et al. (2018), paper‡	45.1	0.01*	0.01*	35.44*
	Artetxe et al. (2018b)	48.13	48.19	32.63	37.33

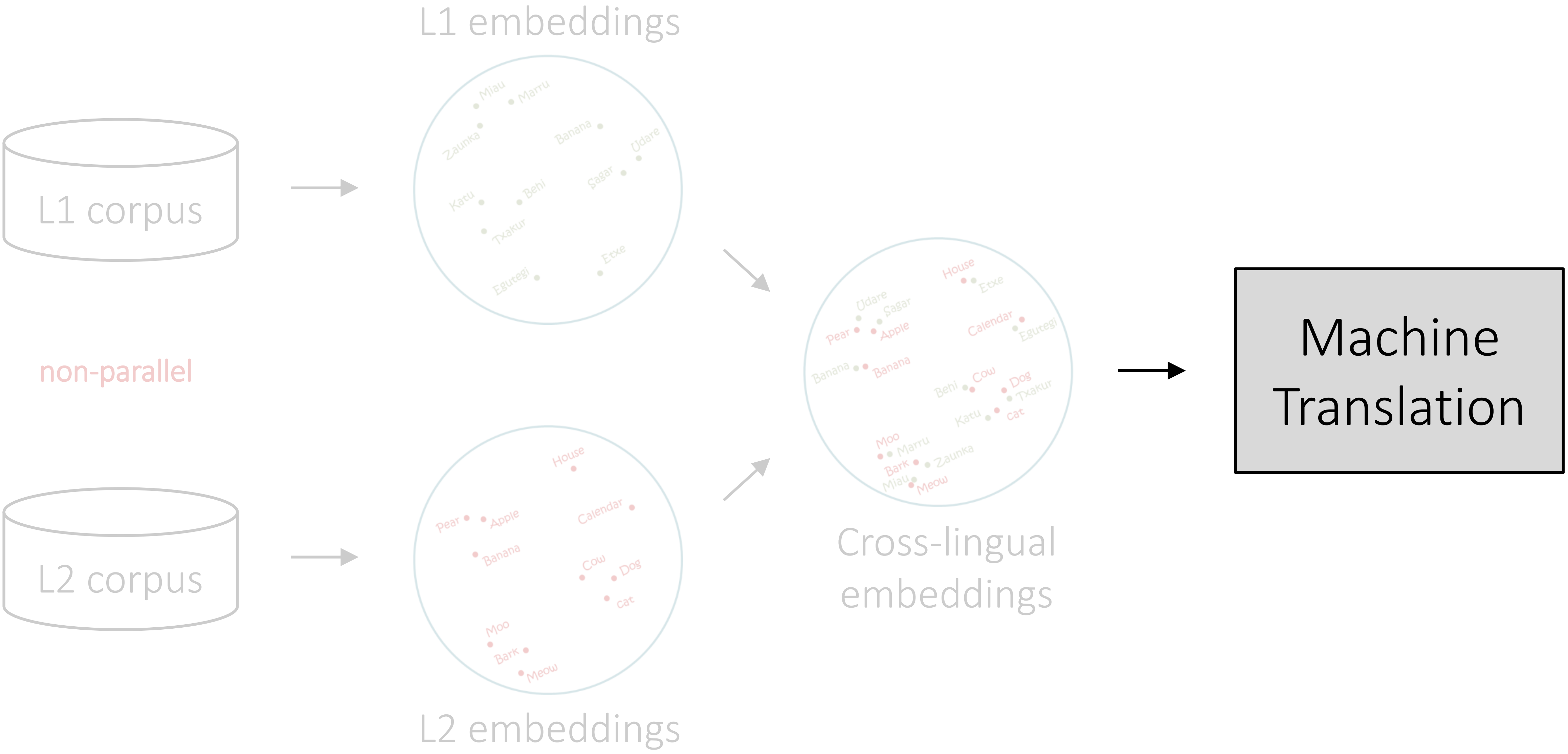
Cross-lingual embedding mappings

Supervision	Method	EN-IT	EN-DE	EN-FI	EN-ES
5k dict.	Mikolov et al. (2013)	34.93†	35.00†	25.91†	27.73†
	Faruqui and Dyer (2014)	38.40*	37.13*	27.60*	26.80*
	Shigeto et al. (2015)	41.53†	43.07†	31.04†	33.73†
	Dinu et al. (2015)	37.7	38.93*	29.14*	30.40*
	Lazaridou et al. (2015)	40.2	-	-	-
	Xing et al. (2015)	36.87†	41.27†	28.23†	31.20†
	Zhang et al. (2016)	36.73†	40.80†	28.16†	31.07†
	Artetxe et al. (2016)	39.27	41.87*	30.62*	31.40*
	Smith et al. (2017)	43.1	43.33†	29.42†	35.13†
	Artetxe et al. (2018a)	45.27	44.13	32.94	36.60
25 dict.	**Artetxe et al. (2017)**	37.27	39.60	28.16	-
Init. heurist.	Smith et al. (2017), cognates	39.9	-	-	-
	Artetxe et al. (2017), num.	39.40	40.27	26.47	-
None	Zhang et al. (2017), $\lambda = 1$	0.00*	0.00*	0.00*	0.00*
	Zhang et al. (2017), $\lambda = 10$	0.00*	0.00*	0.01*	0.01*
	Conneau et al. (2018), code‡	45.15*	46.83*	0.38*	35.38*
	Conneau et al. (2018), paper‡	45.1	0.01*	0.01*	35.44*
	Artetxe et al. (2018b)	48.13	48.19	32.63	37.33

Outline


L1 embeddings
L1 corpus
non-parallel
L2 corpus
L2 embeddings
House
Etxe
Udare
Sagar
Pear
Apple
Calendar
Egutegi
Banana
Banana
Behi
Cow
Dog
Txakur
Katu
Cat
Moo
Marru
Bark
Zaunka
Miau
Meow
Cross-lingual embeddings
Machine Translation

Outline

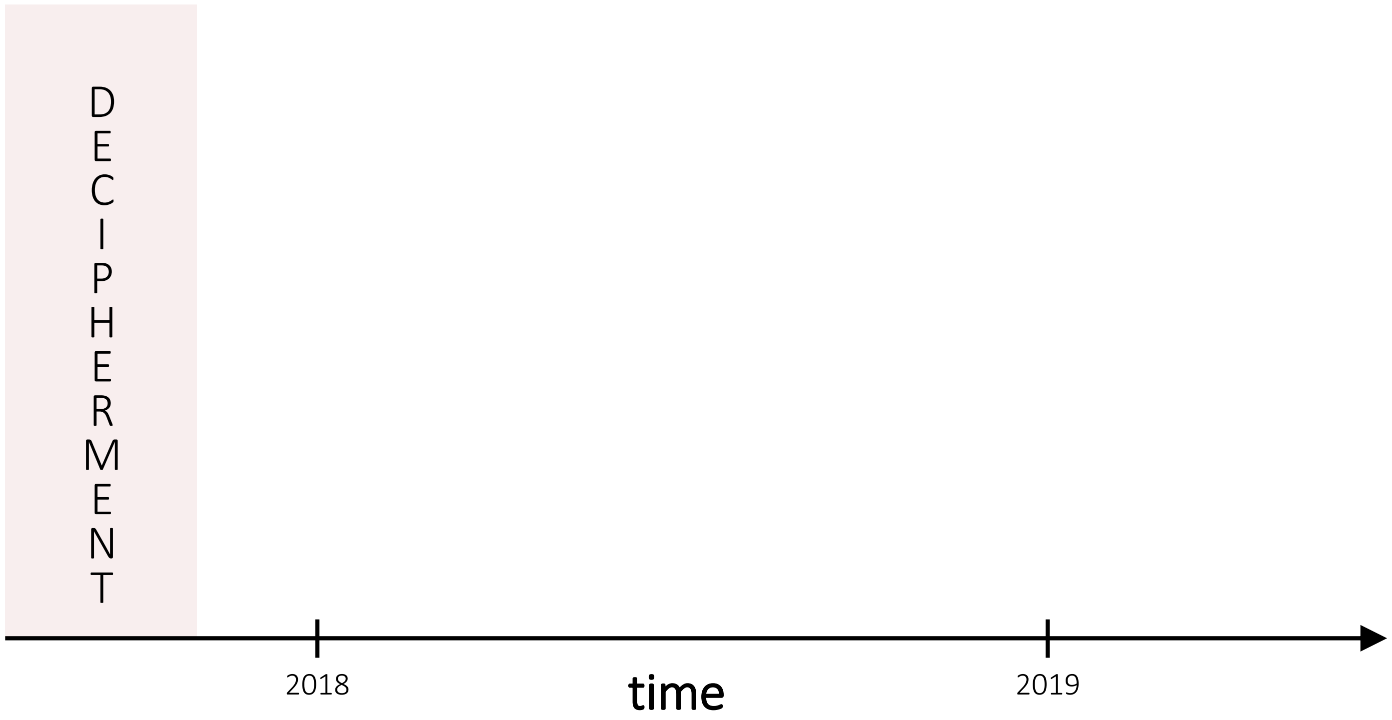


L1 corpus
L1 embeddings
non-parallel
L2 corpus
L2 embeddings
Cross-lingual embeddings
Machine Translation

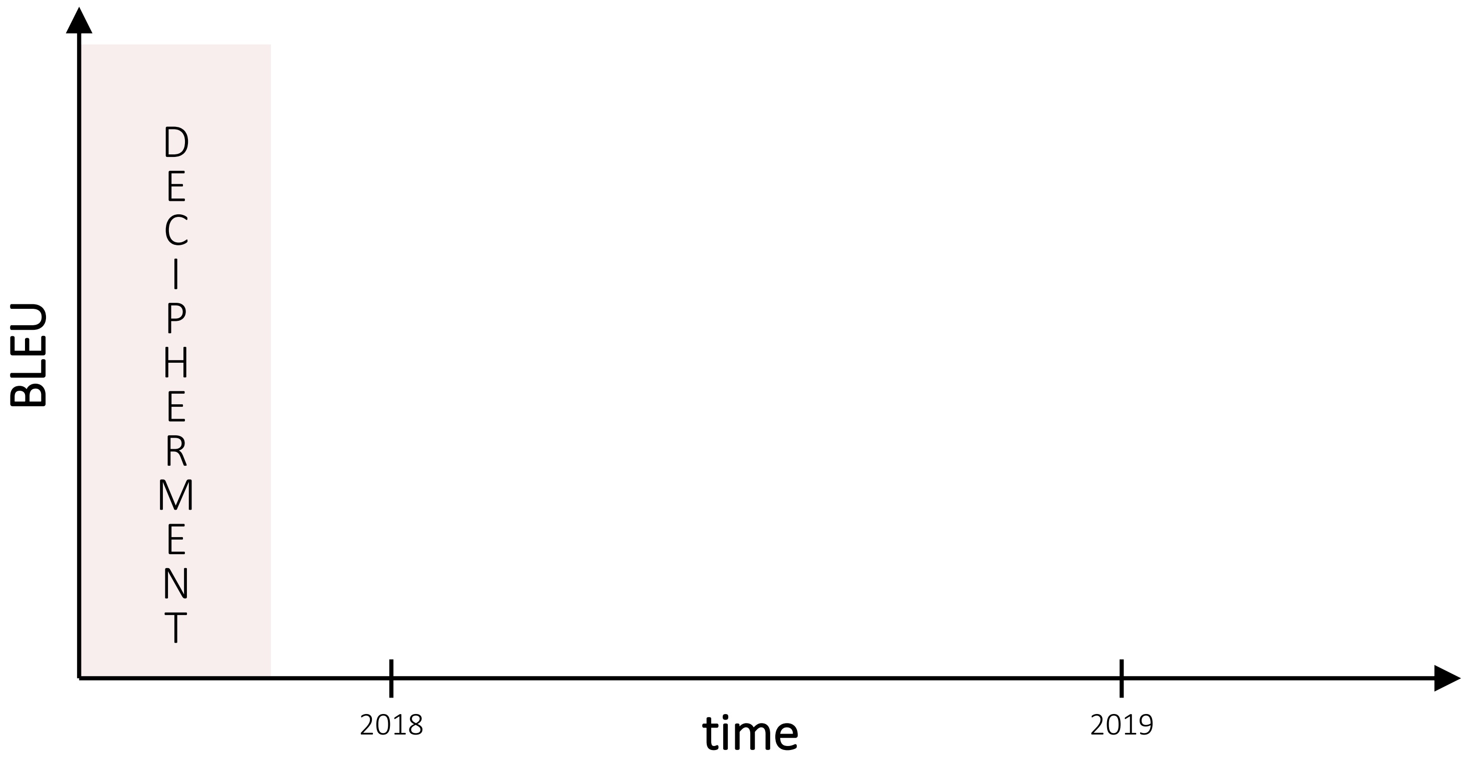
Outline

Outline

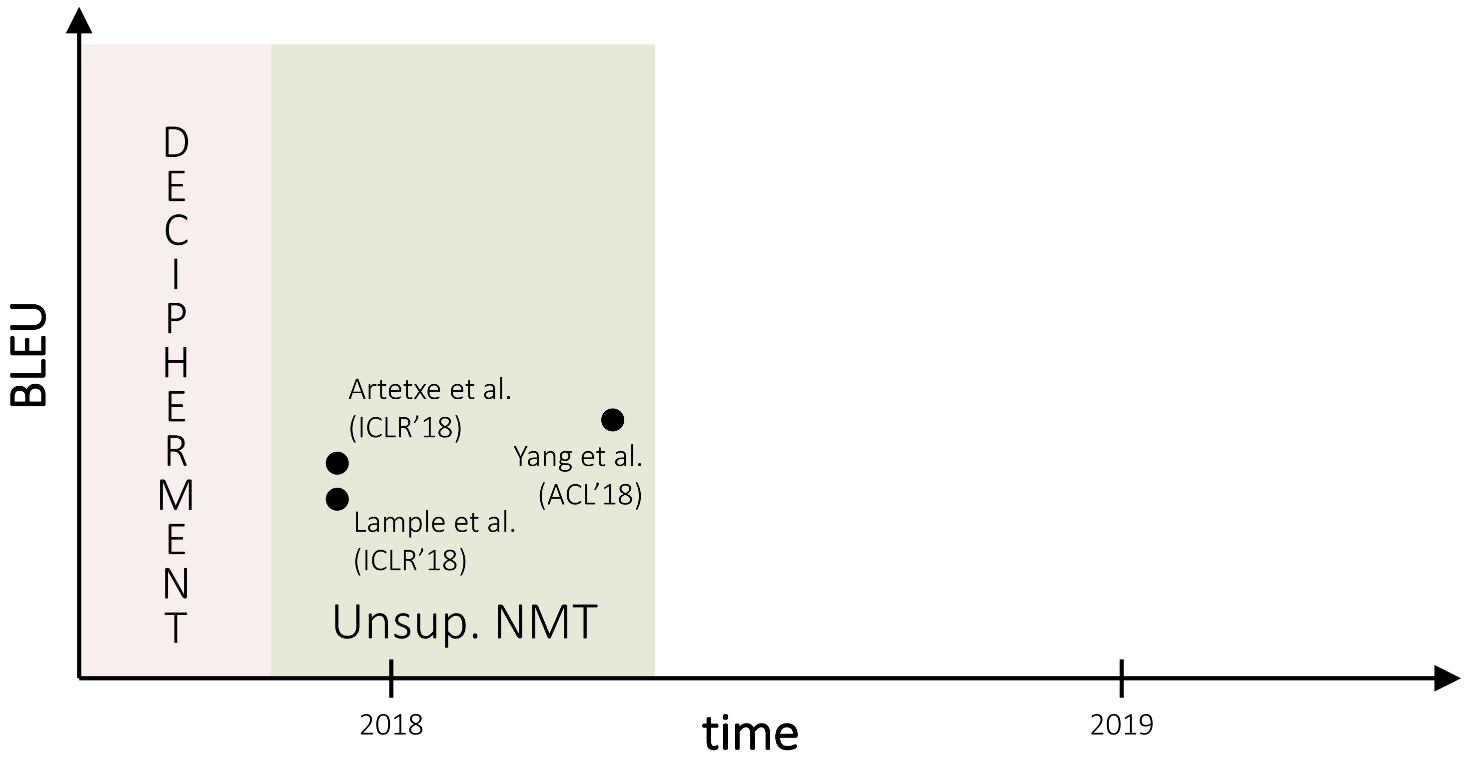
Outline



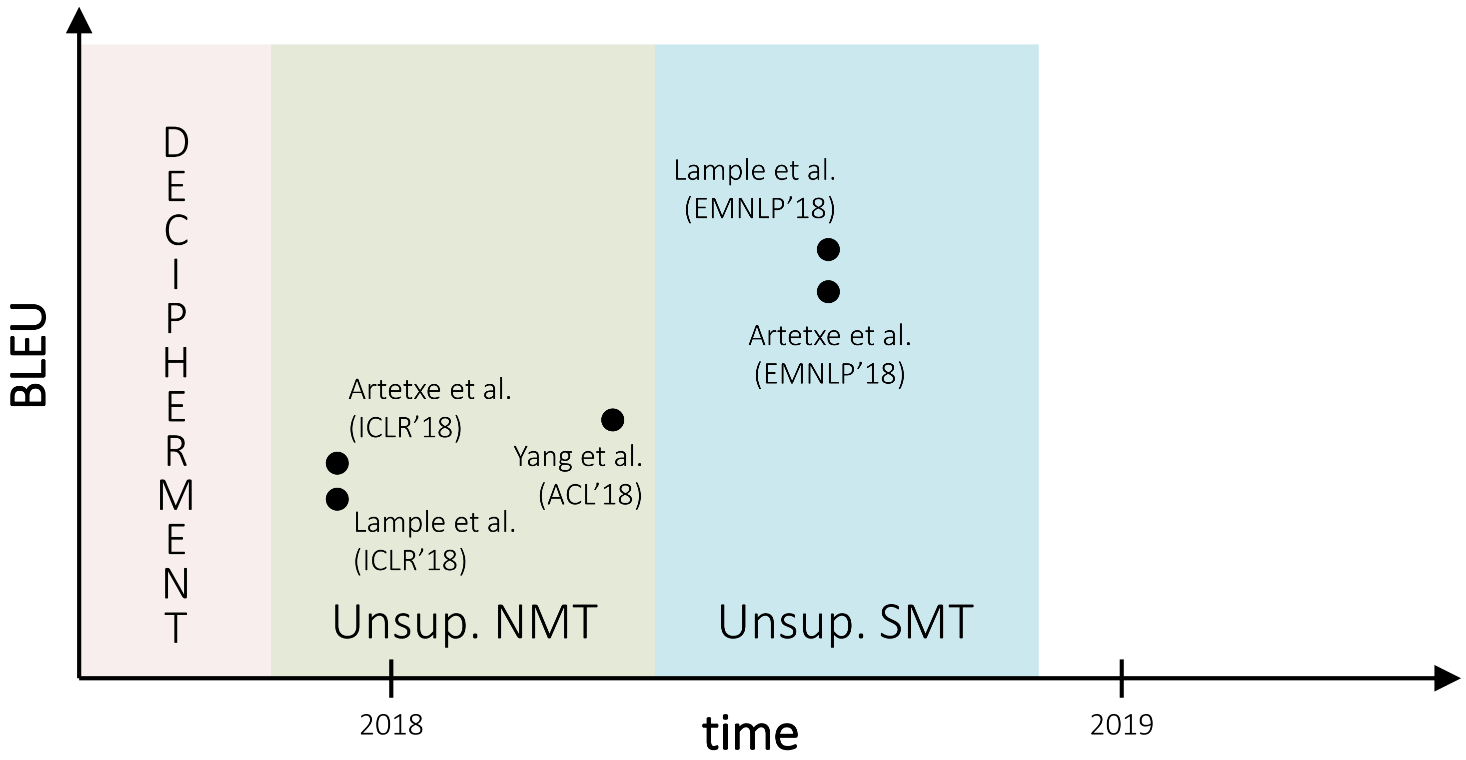
Outline



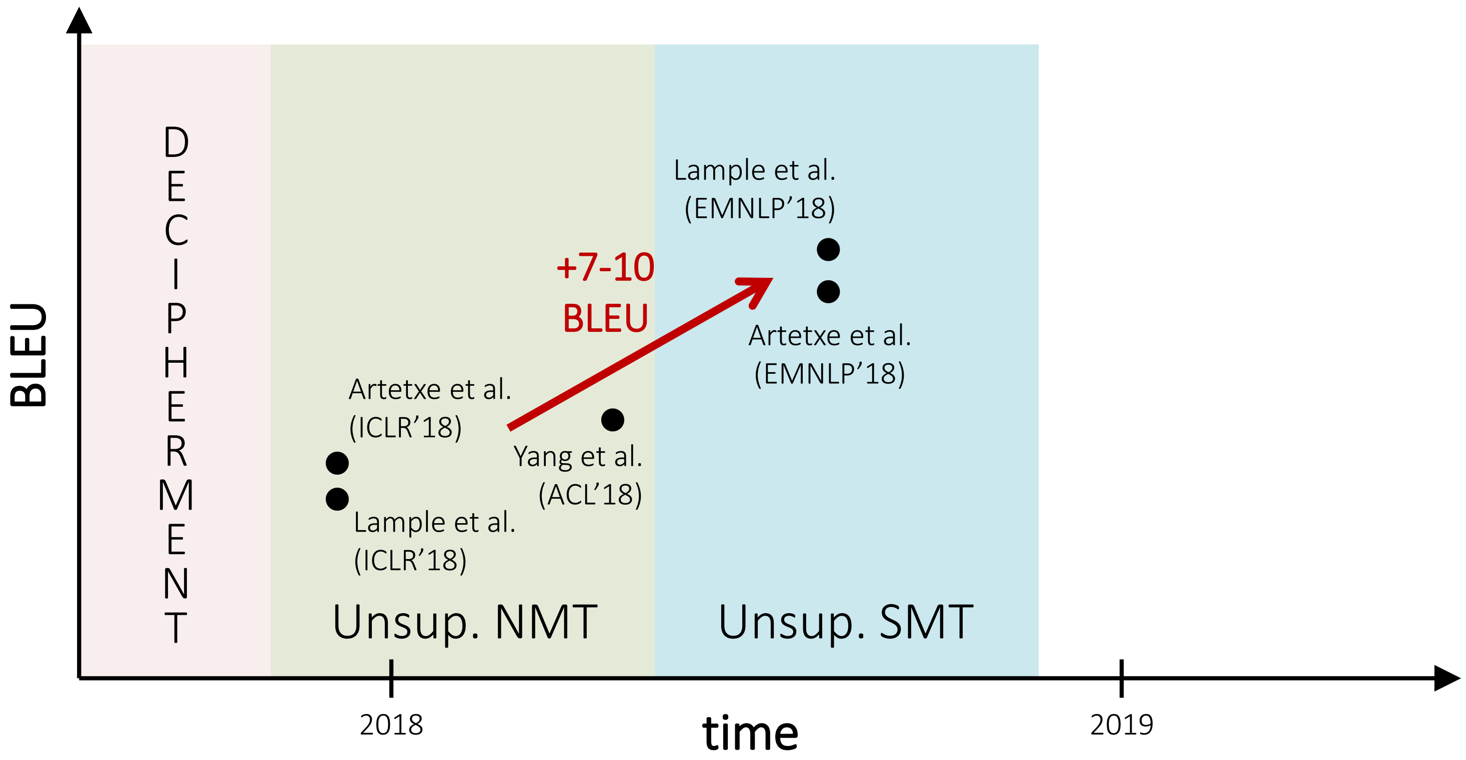
Outline


BLEU
DECIPHERMENT
Artetxe et al. (ICLR'18)
Lample et al. (ICLR'18)
Yang et al. (ACL'18)
Unsup. NMT
2018
time
2019

Outline

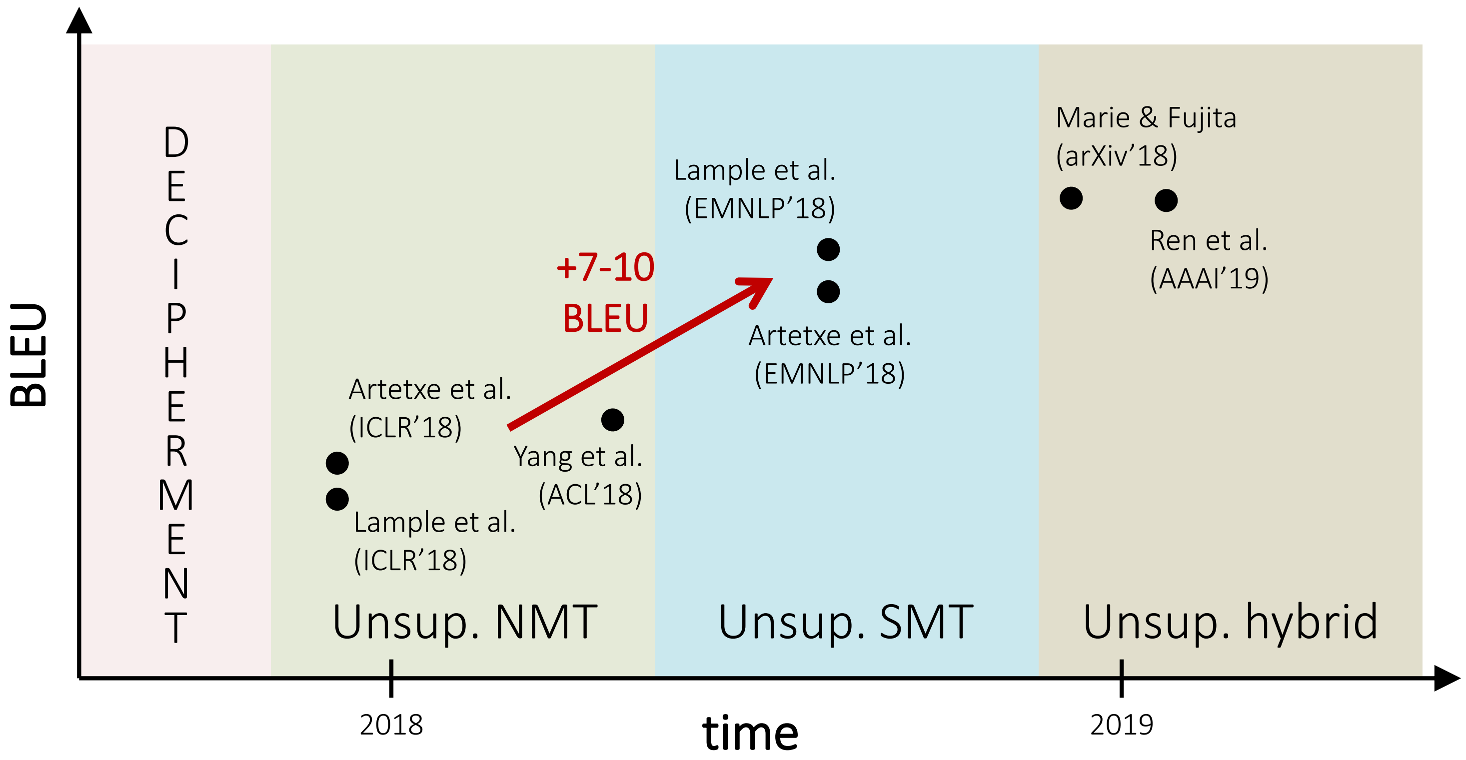

BLEU
DECIPHERMENT
Artetxe et al. (ICLR'18)
Lample et al. (ICLR'18)
Yang et al. (ACL'18)
Unsup. NMT
Lample et al. (EMNLP'18)
Artetxe et al. (EMNLP'18)
Unsup. SMT
2018
time
2019

Outline



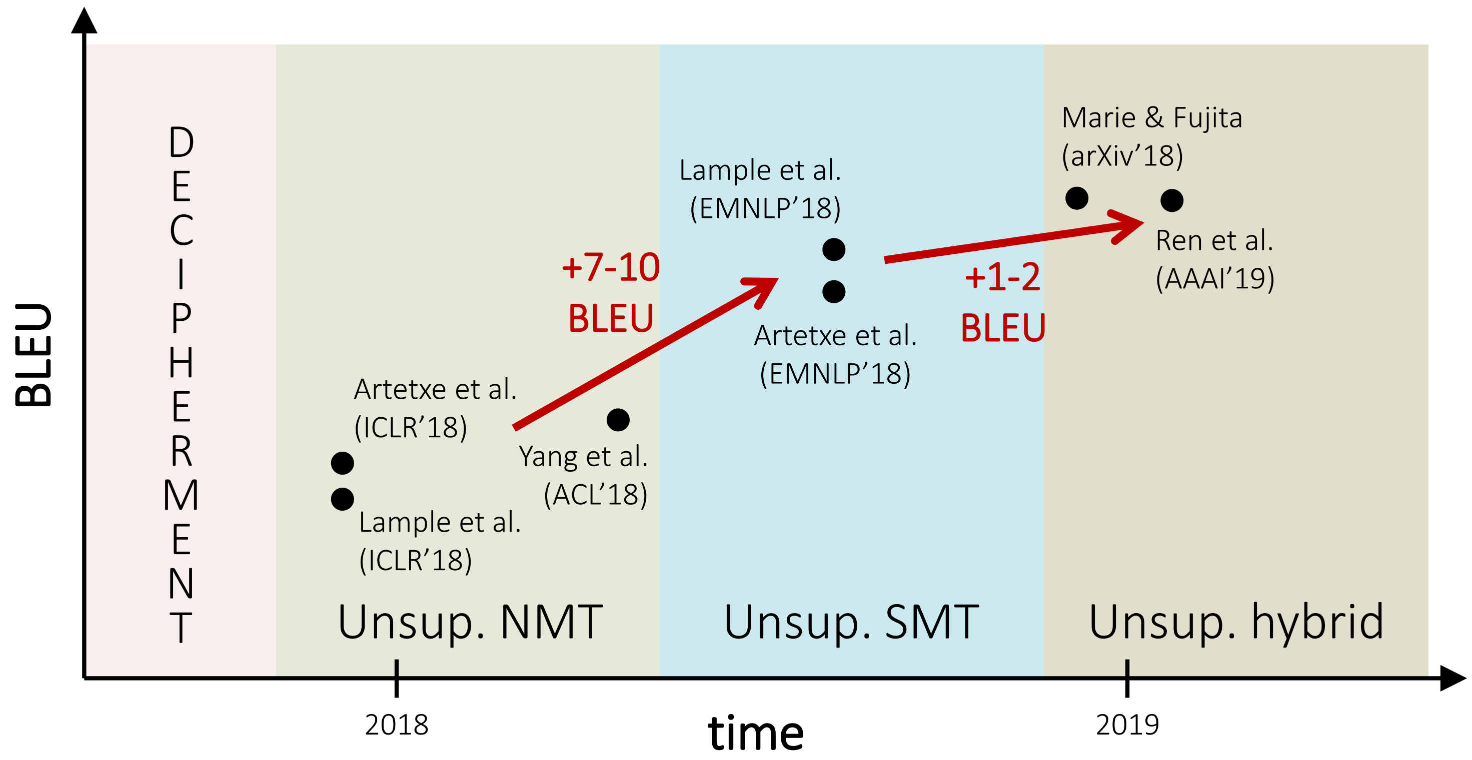
BLEU
DECIPHERMENT
Artetxe et al. (ICLR'18)
Lample et al. (ICLR'18)
Yang et al. (ACL'18)
+7-10 BLEU
Lample et al. (EMNLP'18)
Artetxe et al. (EMNLP'18)
Unsup. NMT
Unsup. SMT
2018
time
2019

Outline



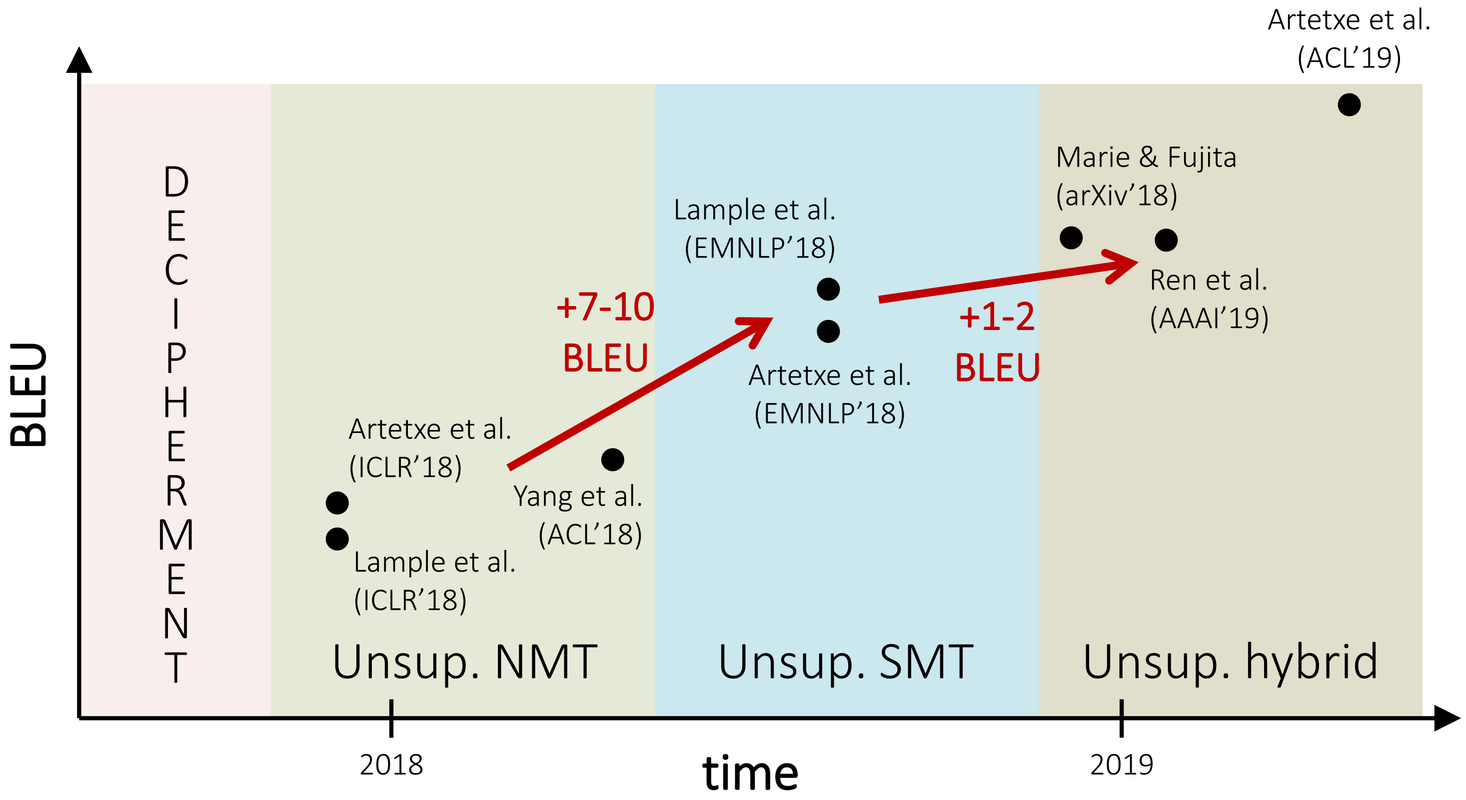
BLEU
D E C I P H E R M E N T
Artetxe et al. (ICLR'18)
Lample et al. (ICLR'18)
Yang et al. (ACL'18)
+7-10 BLEU
Lample et al. (EMNLP'18)
Artetxe et al. (EMNLP'18)
Marie & Fujita (arXiv'18)
Ren et al. (AAAI'19)
Unsup. NMT
Unsup. SMT
Unsup. hybrid
2018
time
2019

Outline



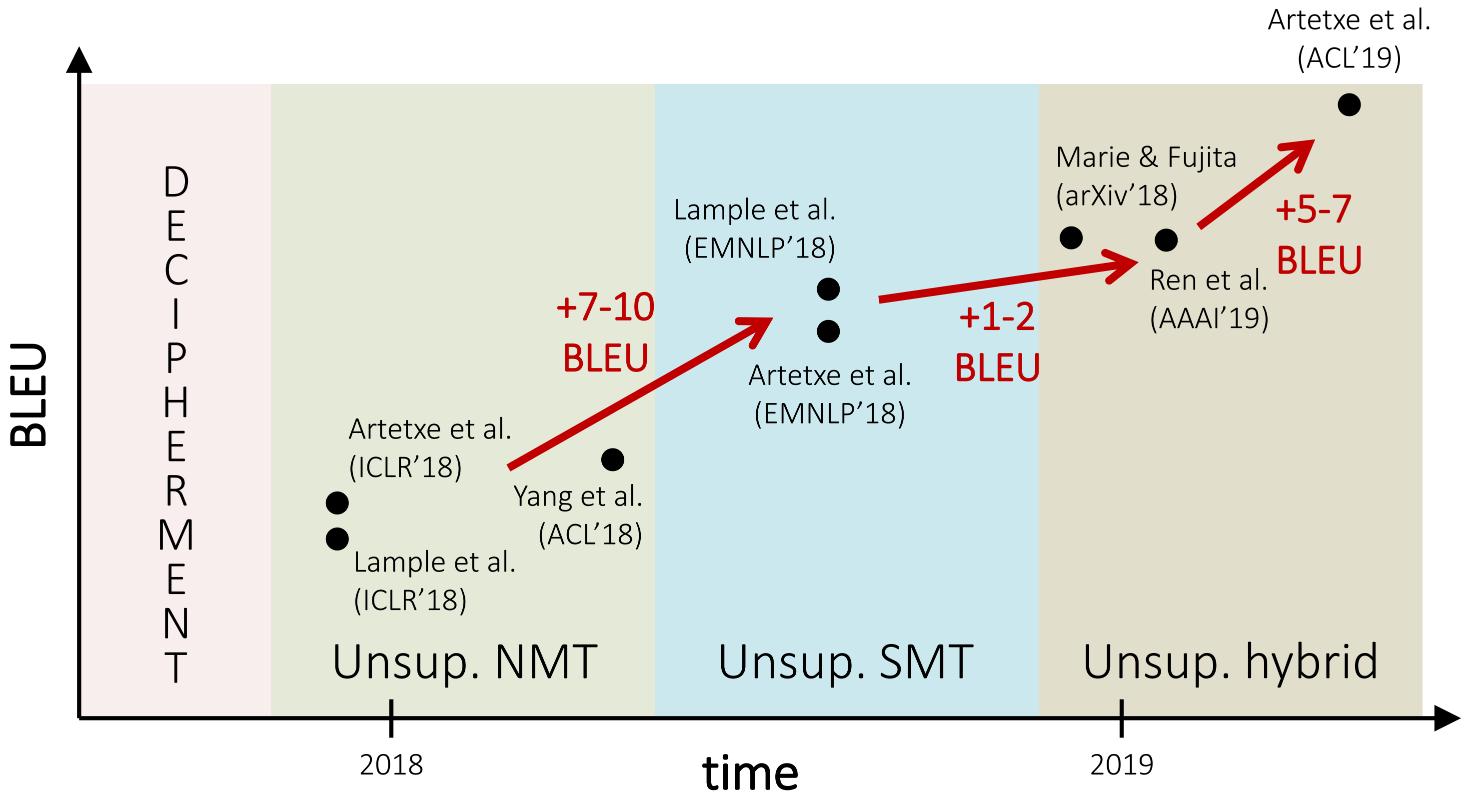
BLEU
DECIPHERMENT
Unsup. NMT
Unsup. SMT
Unsup. hybrid
Artetxe et al. (ICLR'18)
Lample et al. (ICLR'18)
Yang et al. (ACL'18)
+7-10 BLEU
Lample et al. (EMNLP'18)
Artetxe et al. (EMNLP'18)
+1-2 BLEU
Marie & Fujita (arXiv'18)
Ren et al. (AAAI'19)
2018
time
2019

Outline

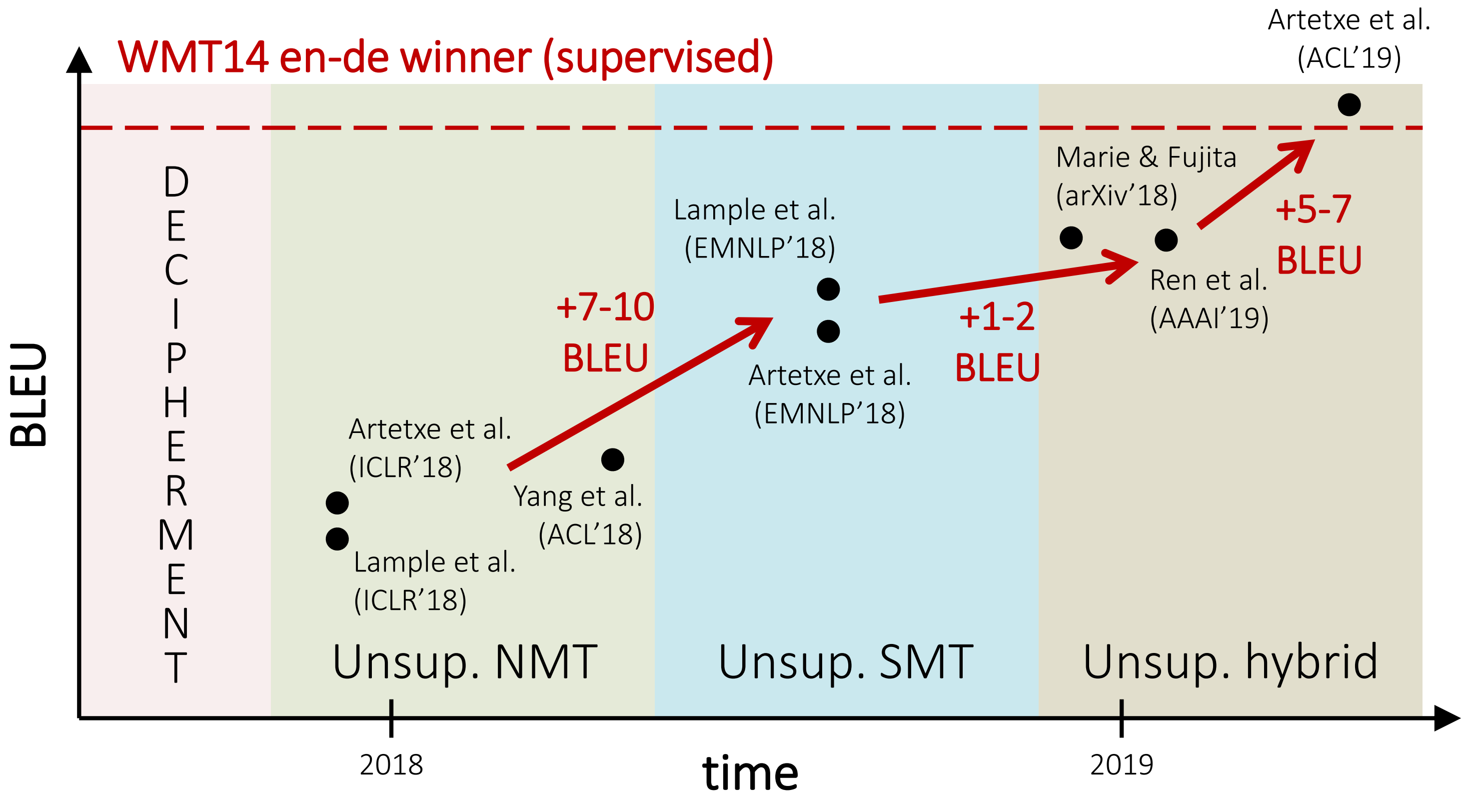


BLEU
DECIPHERMENT
Unsup. NMT
Unsup. SMT
Unsup. hybrid
Artetxe et al. (ICLR'18)
Lample et al. (ICLR'18)
Yang et al. (ACL'18)
+7-10 BLEU
Lample et al. (EMNLP'18)
Artetxe et al. (EMNLP'18)
+1-2 BLEU
Marie & Fujita (arXiv'18)
Ren et al. (AAAI'19)
Artetxe et al. (ACL'19)
2018
time
2019

Outline

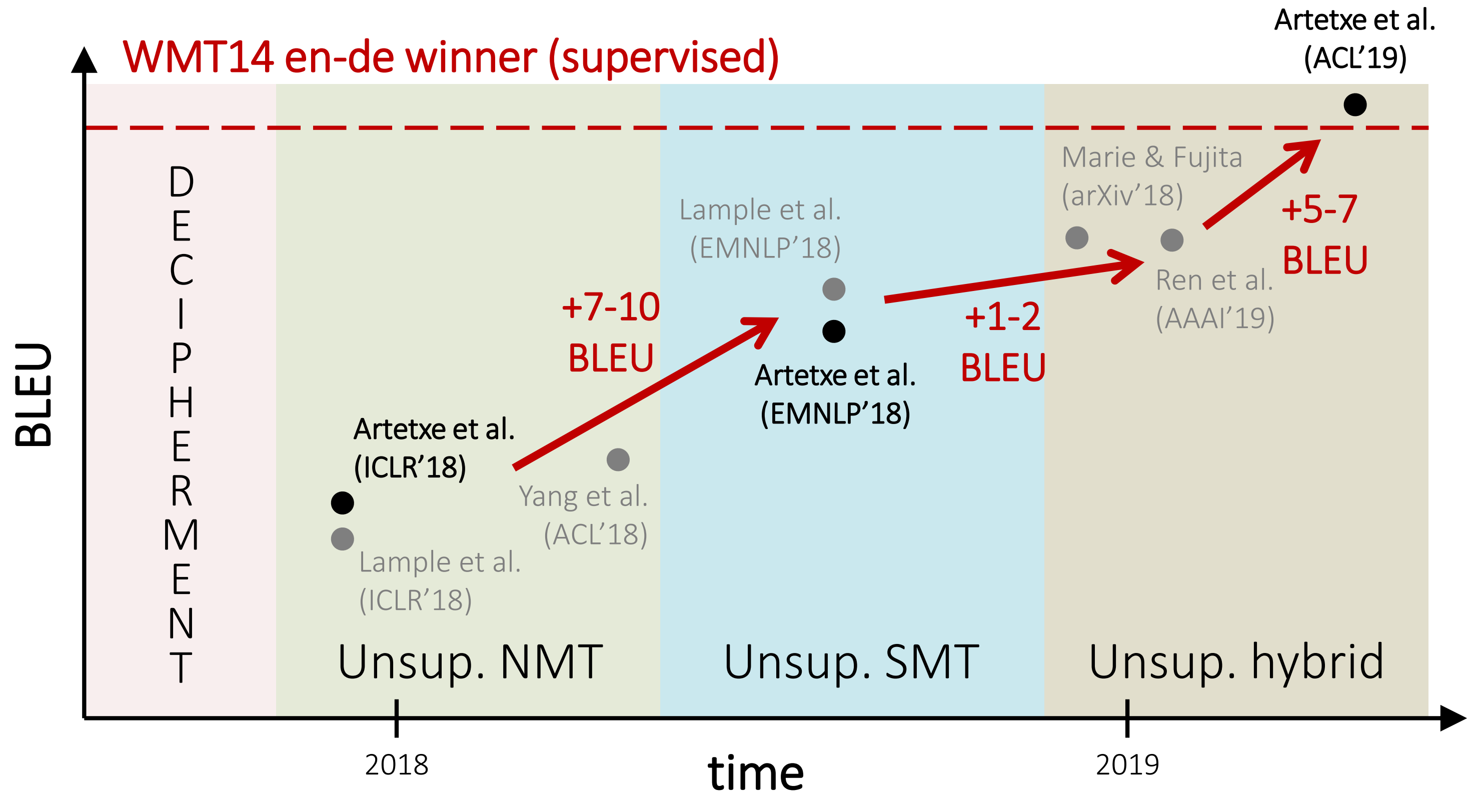

Artetxe et al.
(ACL'19)
BLEU
DECIPHERMENT
Marie & Fujita
(arXiv'18)
+5-7
BLEU
Lample et al.
(EMNLP'18)
Ren et al.
(AAAI'19)
+7-10
BLEU
+1-2
BLEU
Artetxe et al.
(EMNLP'18)
Artetxe et al.
(ICLR'18)
Yang et al.
(ACL'18)
Lample et al.
(ICLR'18)
Unsup. NMT
Unsup. SMT
Unsup. hybrid
2018
time
2019

Outline



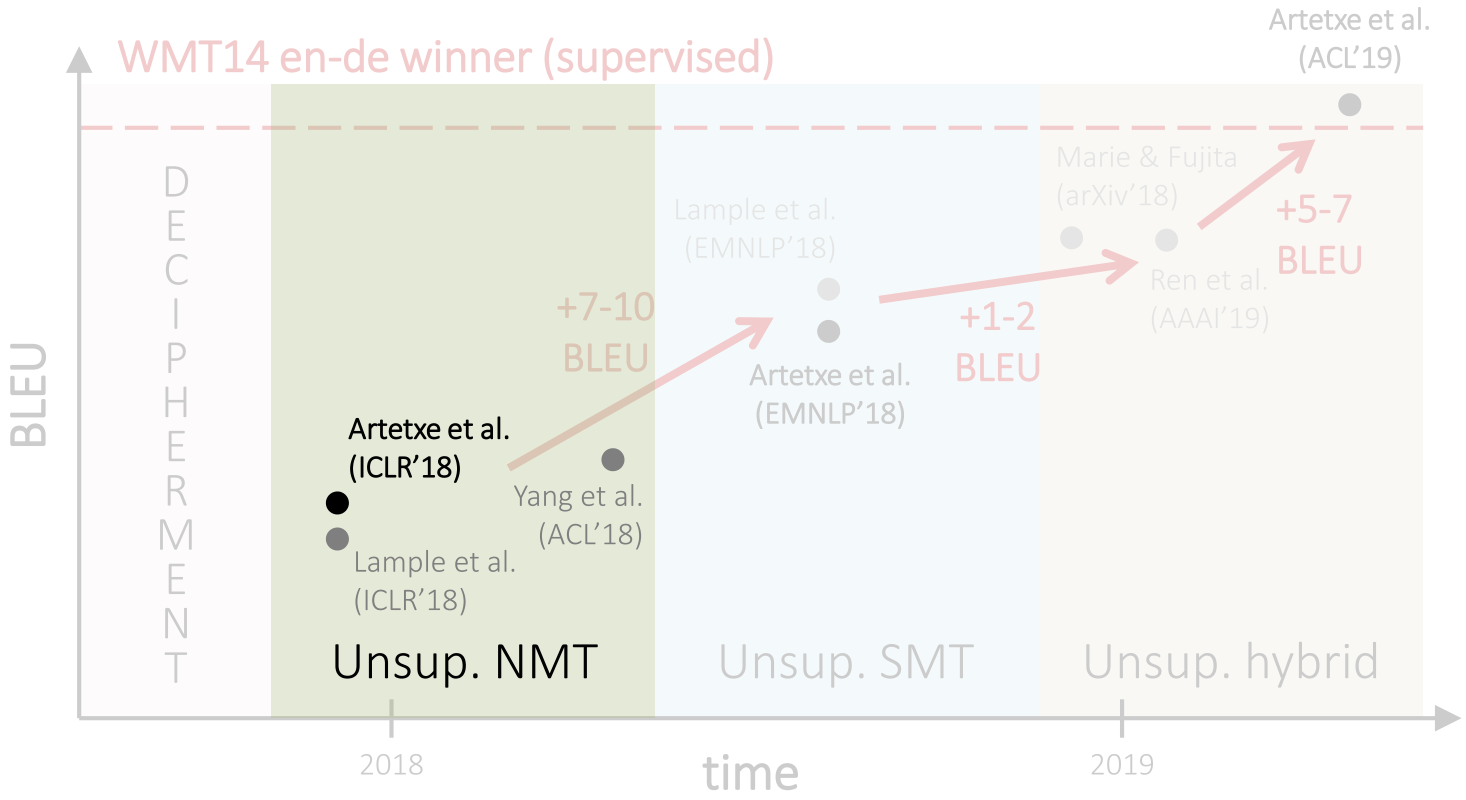
WMT14 en-de winner (supervised)
BLEU
DECIPHERMENT
Artetxe et al. (ICLR'18)
Lample et al. (ICLR'18)
Yang et al. (ACL'18)
+7-10 BLEU
Lample et al. (EMNLP'18)
Artetxe et al. (EMNLP'18)
+1-2 BLEU
Marie & Fujita (arXiv'18)
Ren et al. (AAAI'19)
+5-7 BLEU
Artetxe et al. (ACL'19)
Unsup. NMT
Unsup. SMT
Unsup. hybrid
2018
time
2019

Outline



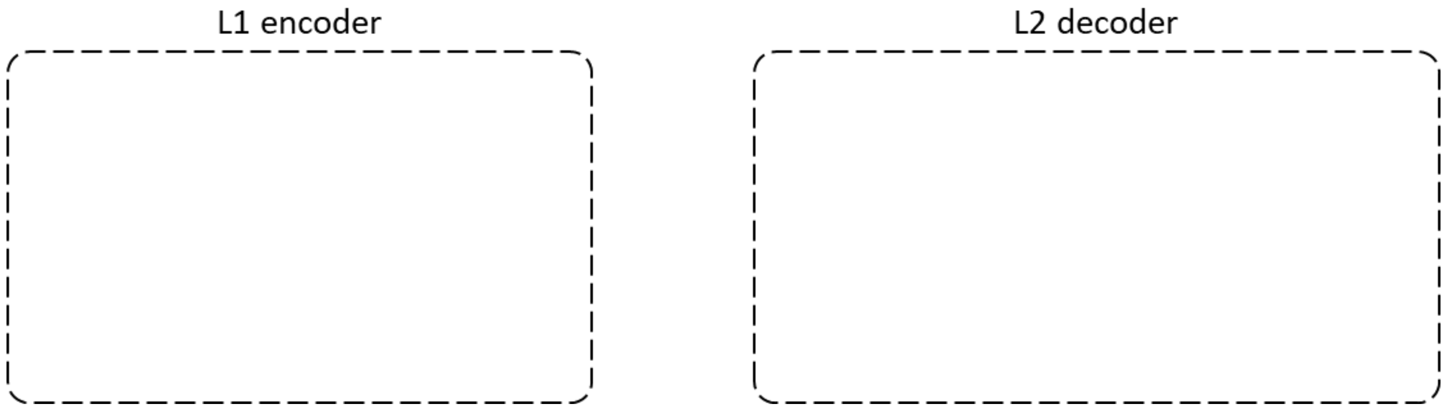
WMT14 en-de winner (supervised)
BLEU
DECIPHERMENT
Artetxe et al.
(ICLR'18)
Lample et al.
(ICLR'18)
Yang et al.
(ACL'18)
+7-10
BLEU
Lample et al.
(EMNLP'18)
Artetxe et al.
(EMNLP'18)
+1-2
BLEU
Marie & Fujita
(arXiv'18)
Ren et al.
(AAAI'19)
+5-7
BLEU
Artetxe et al.
(ACL'19)
Unsup. NMT
Unsup. SMT
Unsup. hybrid
2018
time
2019

Outline

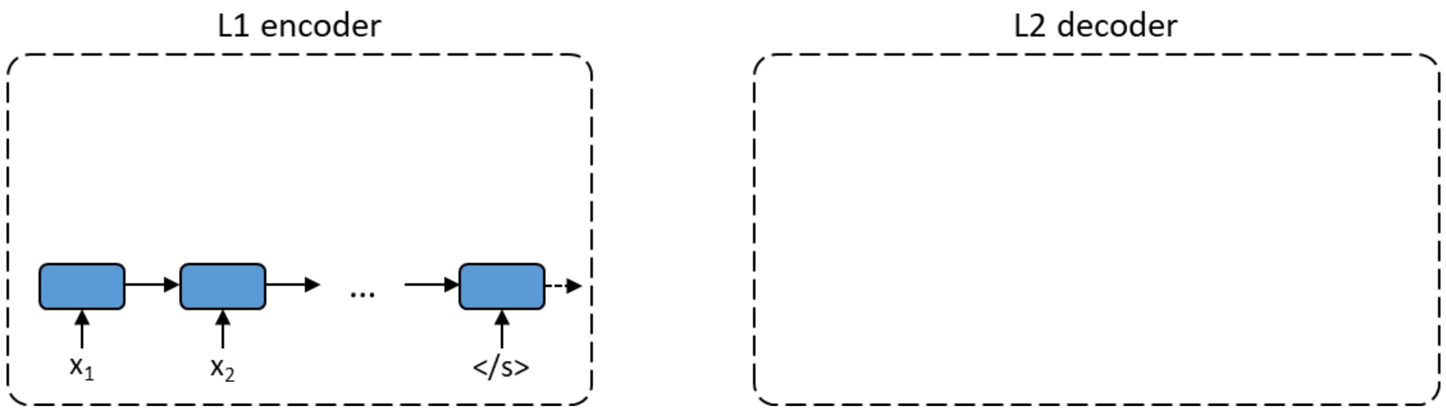

WMT14 en-de winner (supervised)
BLEU
DECIPHERMENT
Unsup. NMT
Unsup. SMT
Unsup. hybrid
Artetxe et al.
(ICLR'18)
Lample et al.
(ICLR'18)
Yang et al.
(ACL'18)
+7-10
BLEU
Lample et al.
(EMNLP'18)
Artetxe et al.
(EMNLP'18)
+1-2
BLEU
Marie & Fujita
(arXiv'18)
Ren et al.
(AAAI'19)
+5-7
BLEU
Artetxe et al.
(ACL'19)
2018
time
2019

Neural machine translation

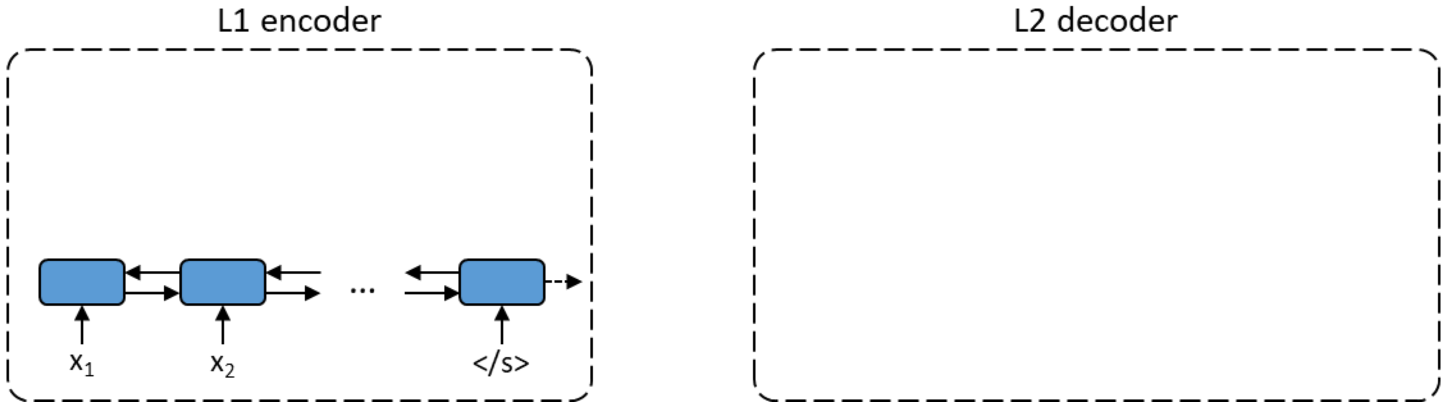
Neural machine translation



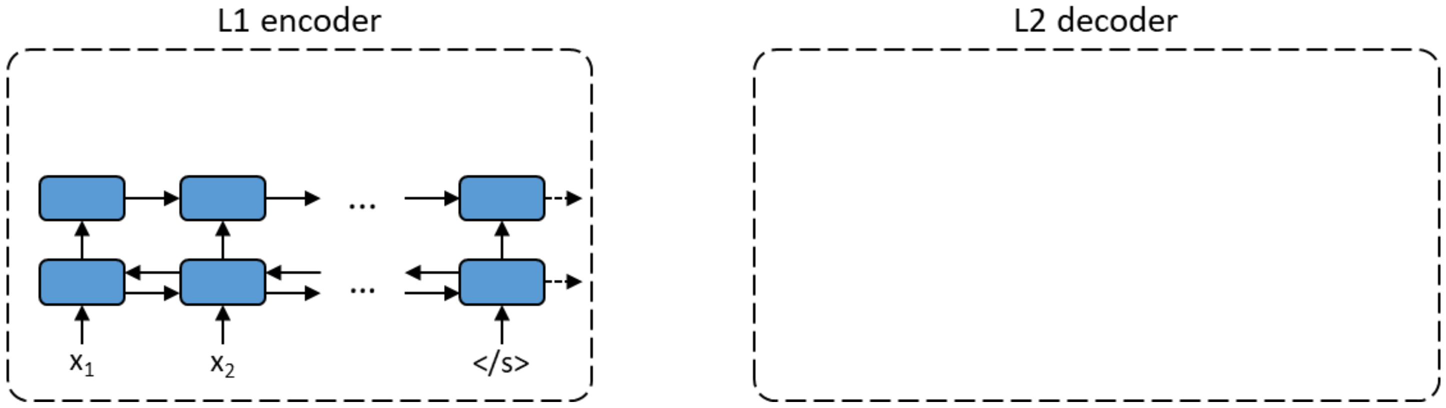
Neural machine translation



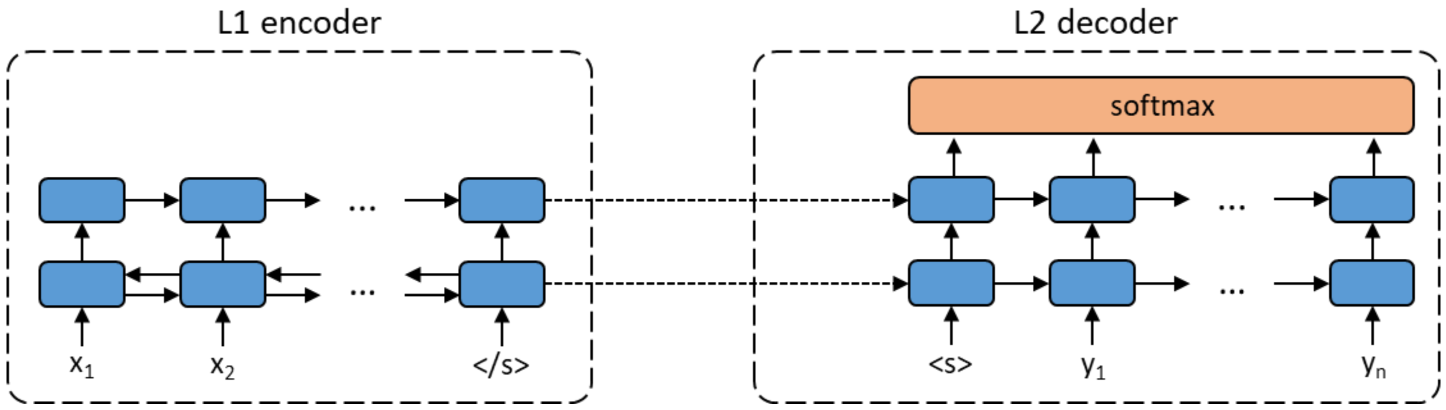
Neural machine translation



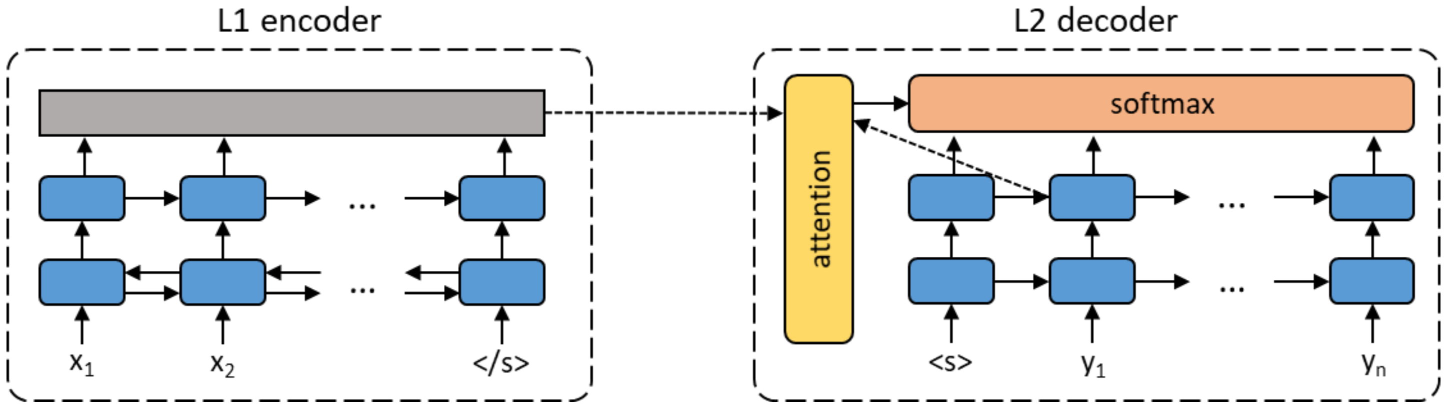
Neural machine translation



Neural machine translation

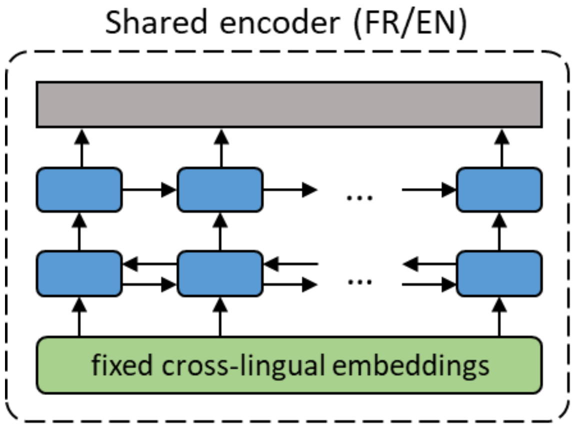


Neural machine translation

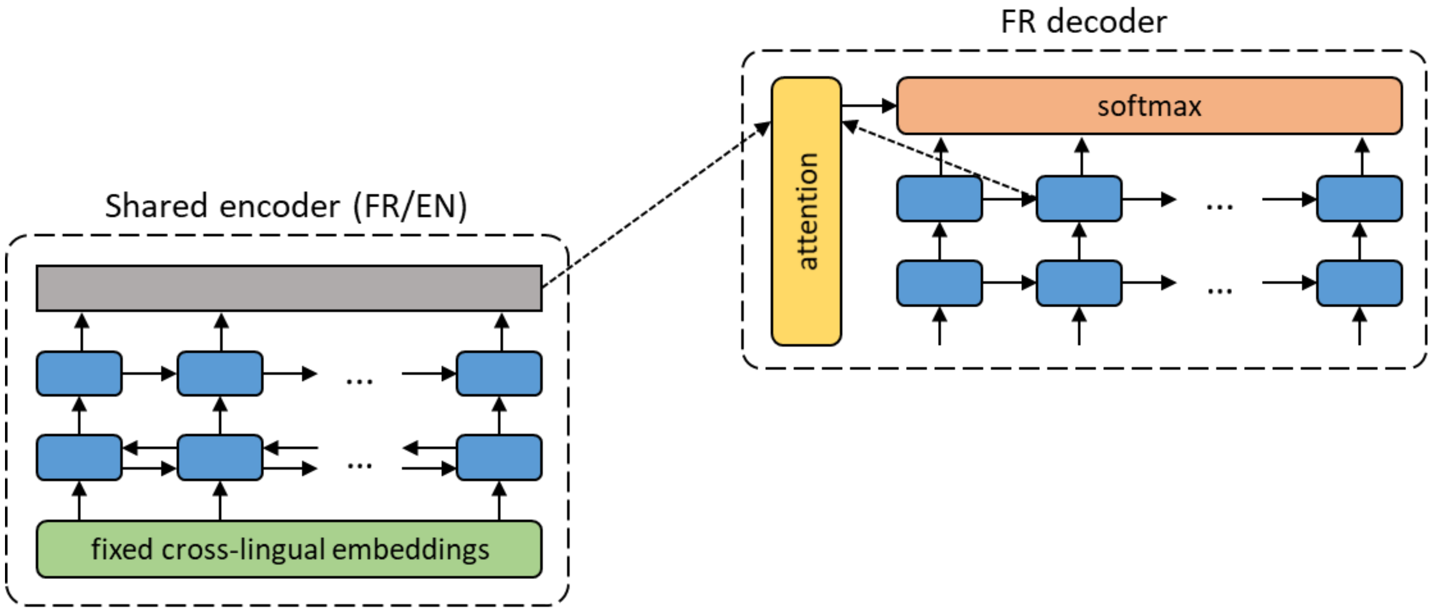


Unsupervised neural machine translation

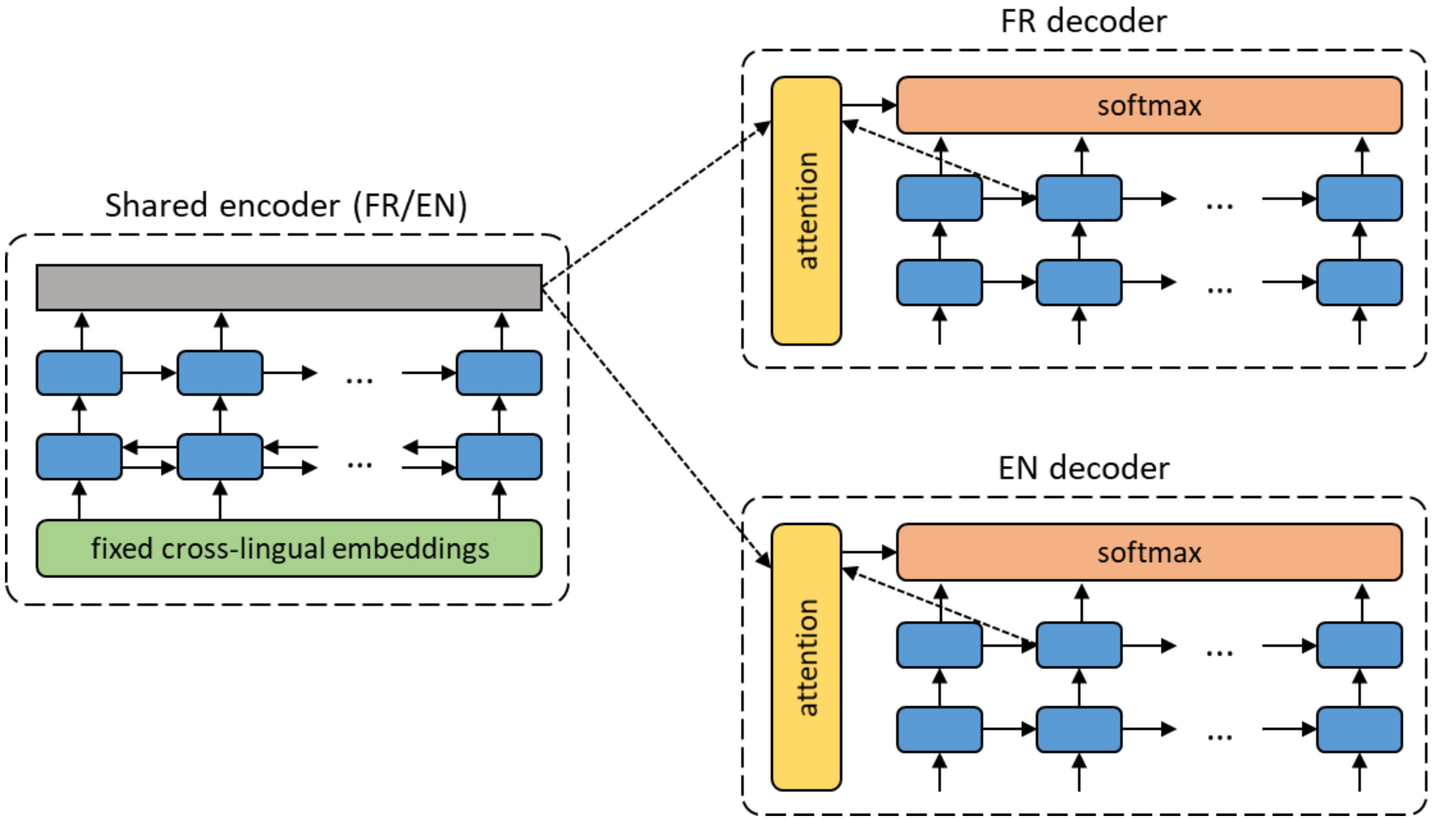
Unsupervised neural machine translation



Unsupervised neural machine translation

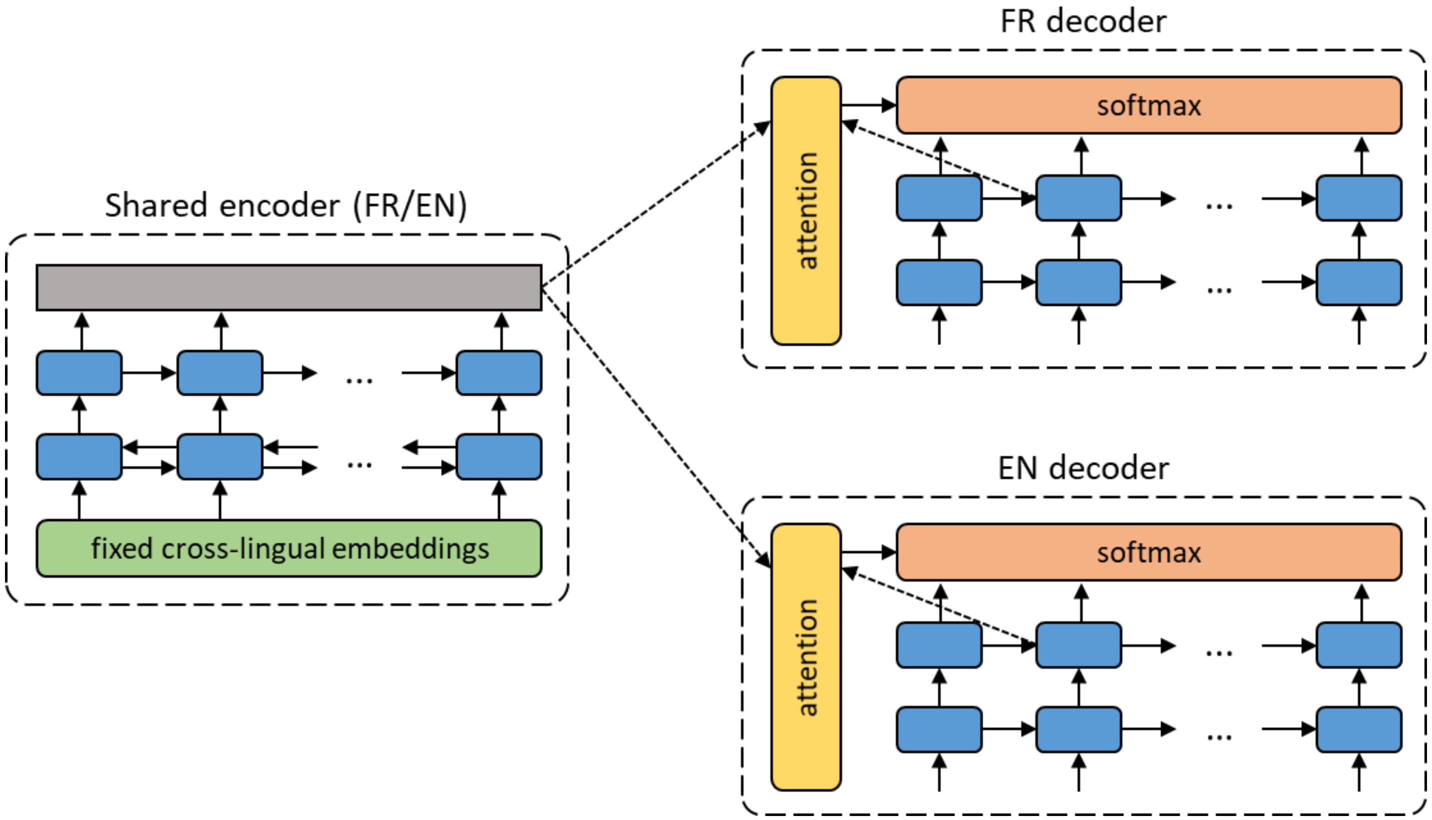


Unsupervised neural machine translation



Unsupervised neural machine translation

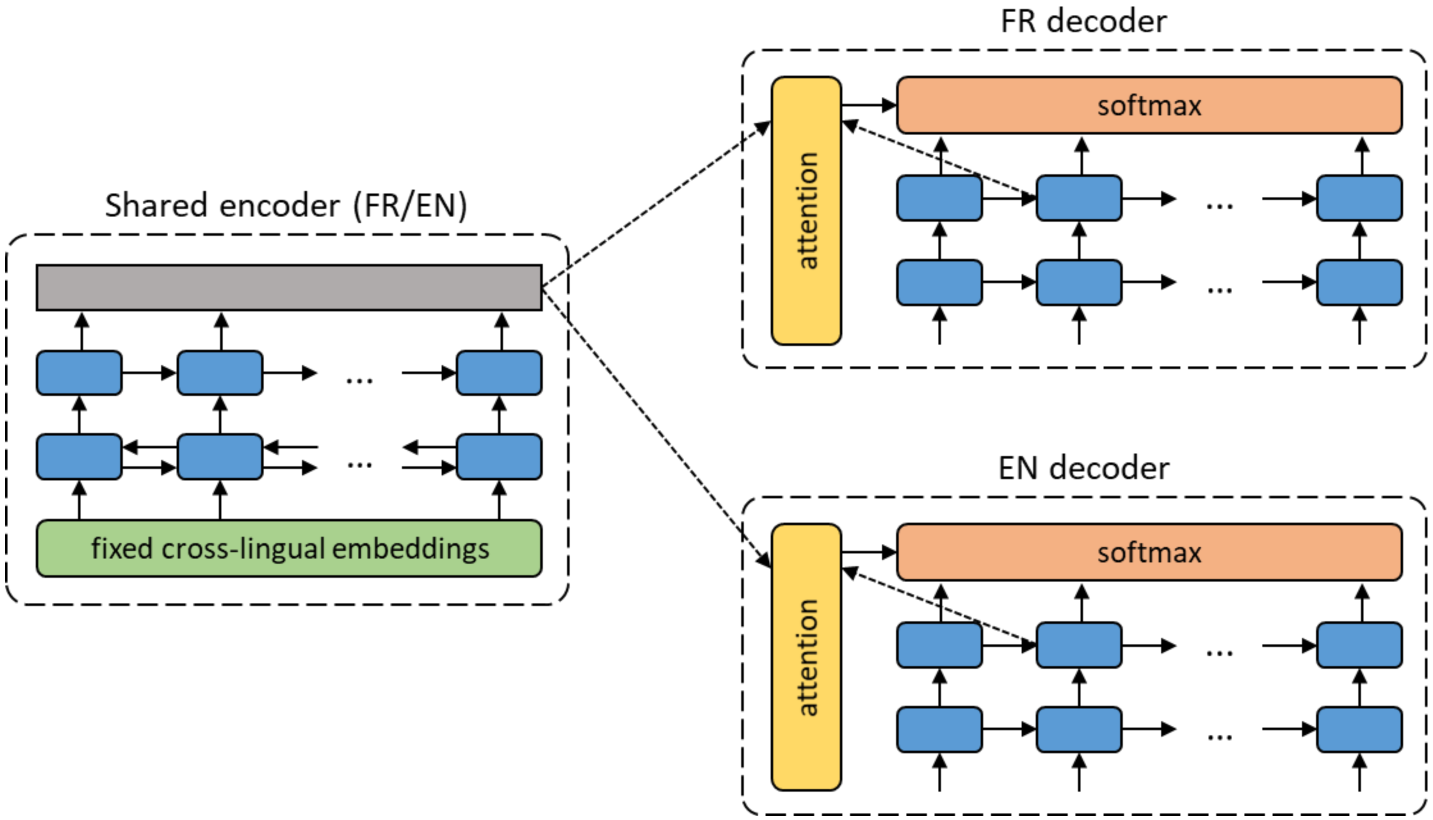
Training



Unsupervised neural machine translation

Training

Une fusillade a eu lieu à l'aéroport international de Los Angeles.

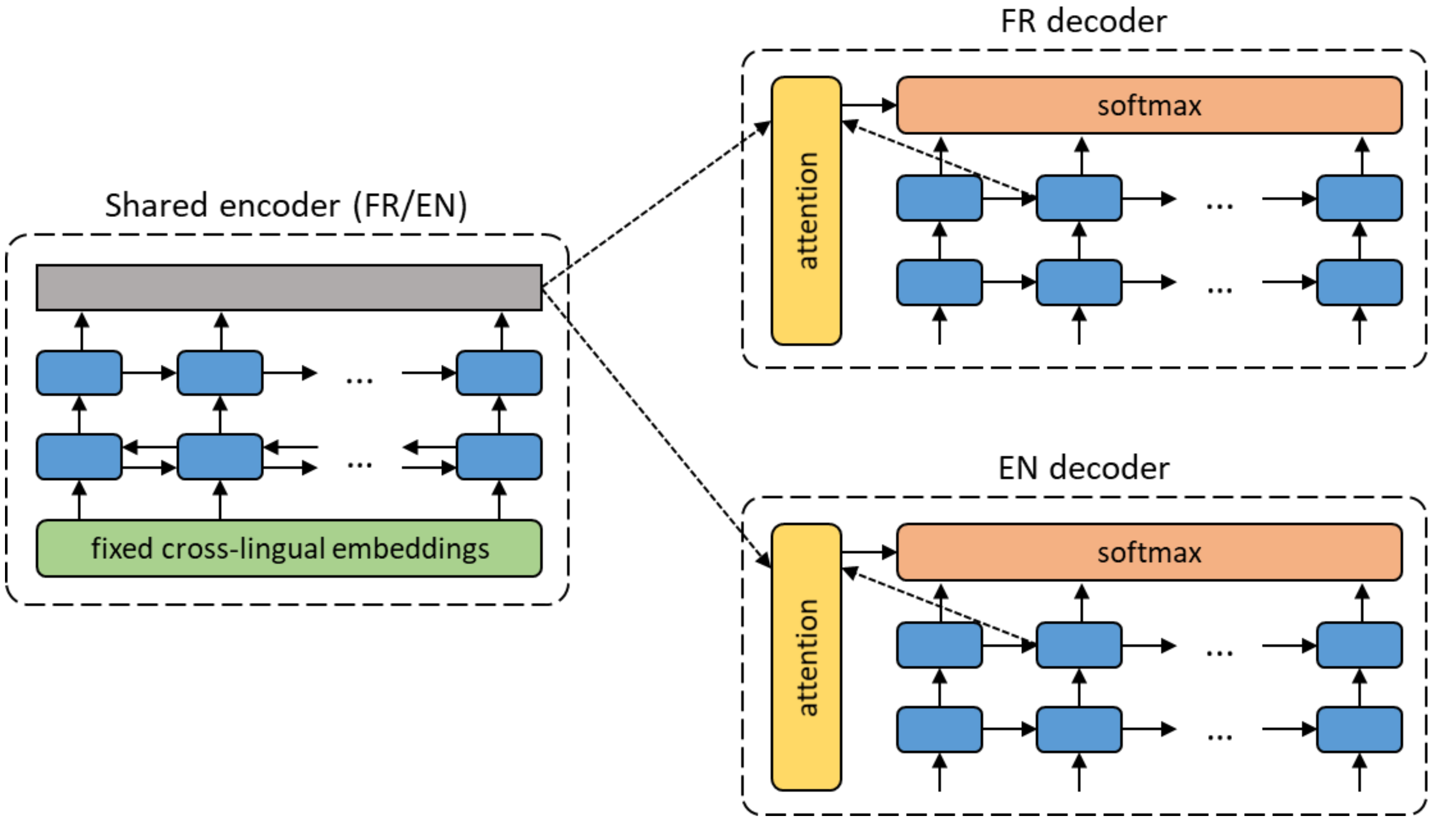


Unsupervised neural machine translation

Training

- Supervised

Une fusillade a eu lieu à l'aéroport international de Los Angeles.

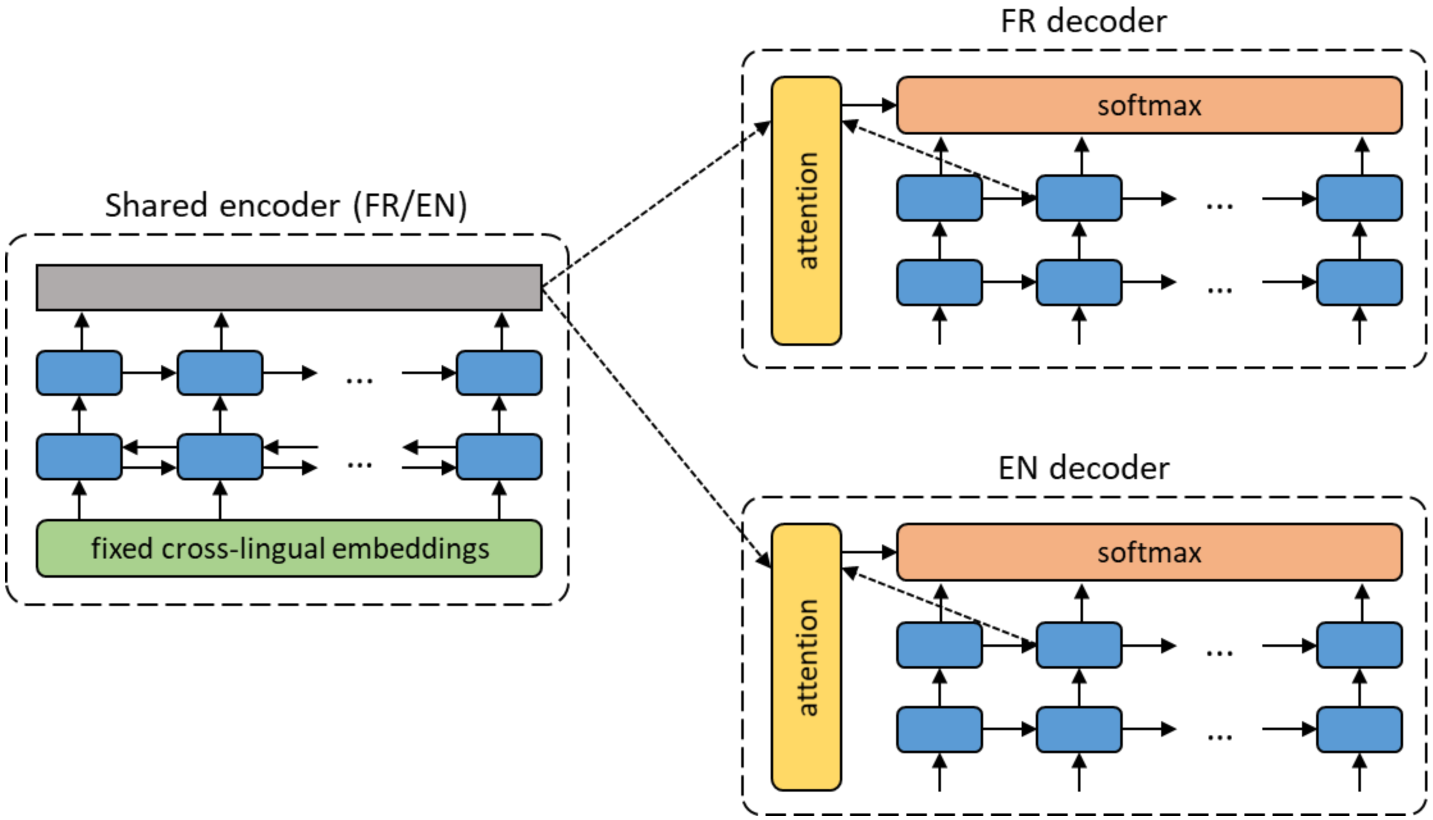


Unsupervised neural machine translation

Training

- Supervised

Une fusillade a eu lieu à l'aéroport international de Los Angeles.



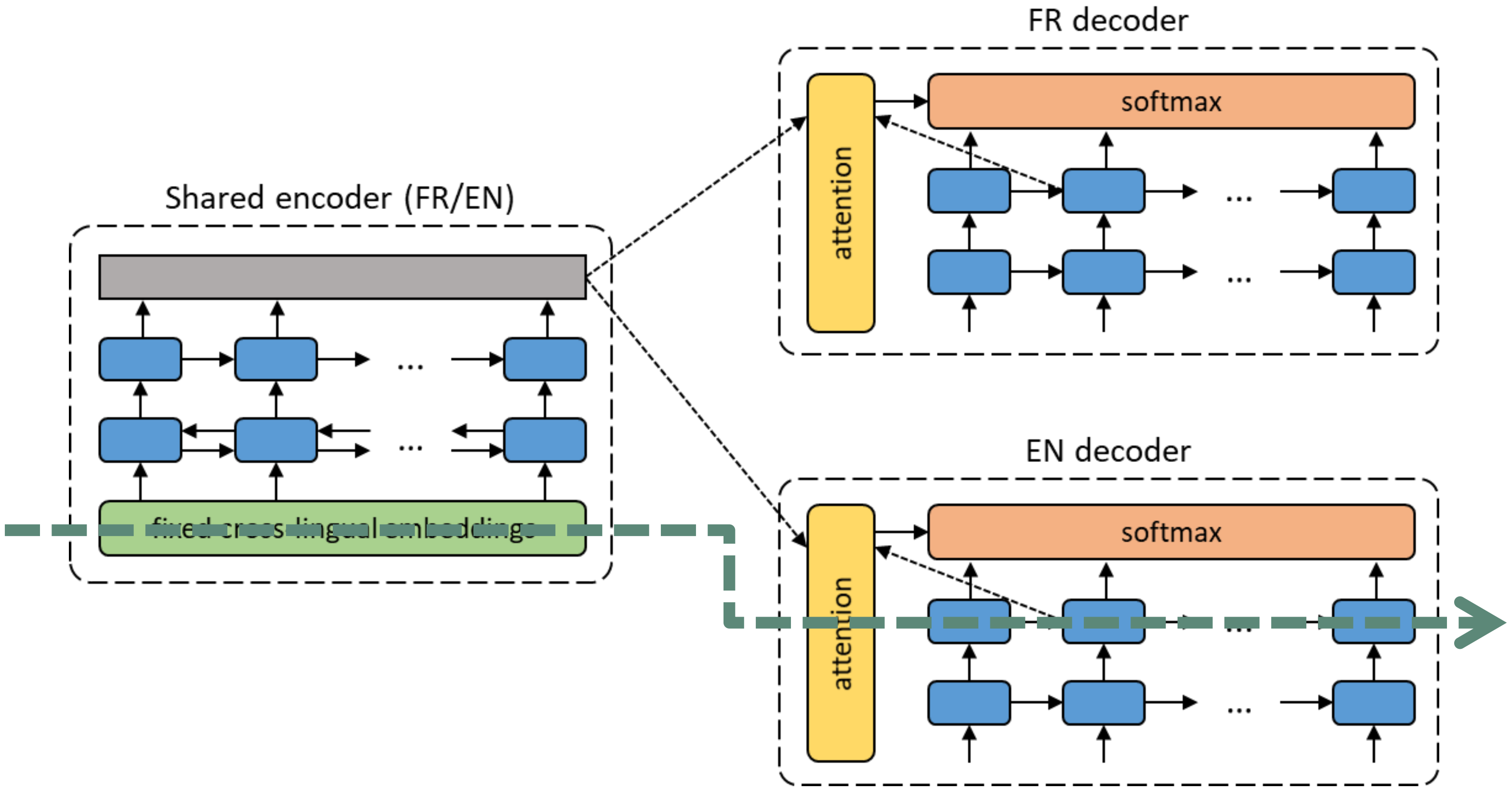
There was a shooting in Los Angeles International Airport.

Unsupervised neural machine translation

Training

- Supervised

Une fusillade a eu lieu à l'aéroport international de Los Angeles.



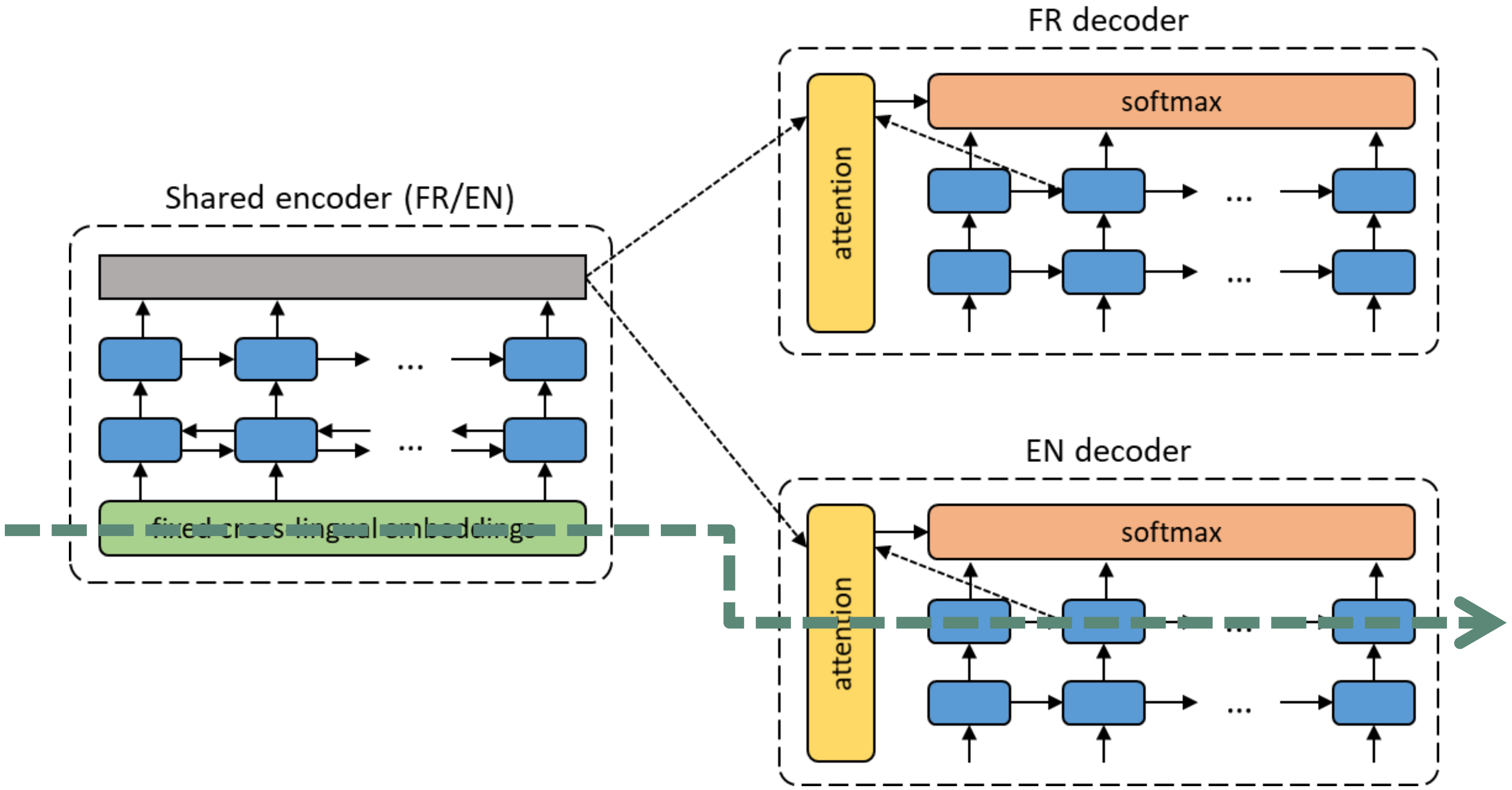
There was a shooting in Los Angeles International Airport.

Unsupervised neural machine translation

Training

- Supervised

Une fusillade a eu lieu à l'aéroport international de Los Angeles.



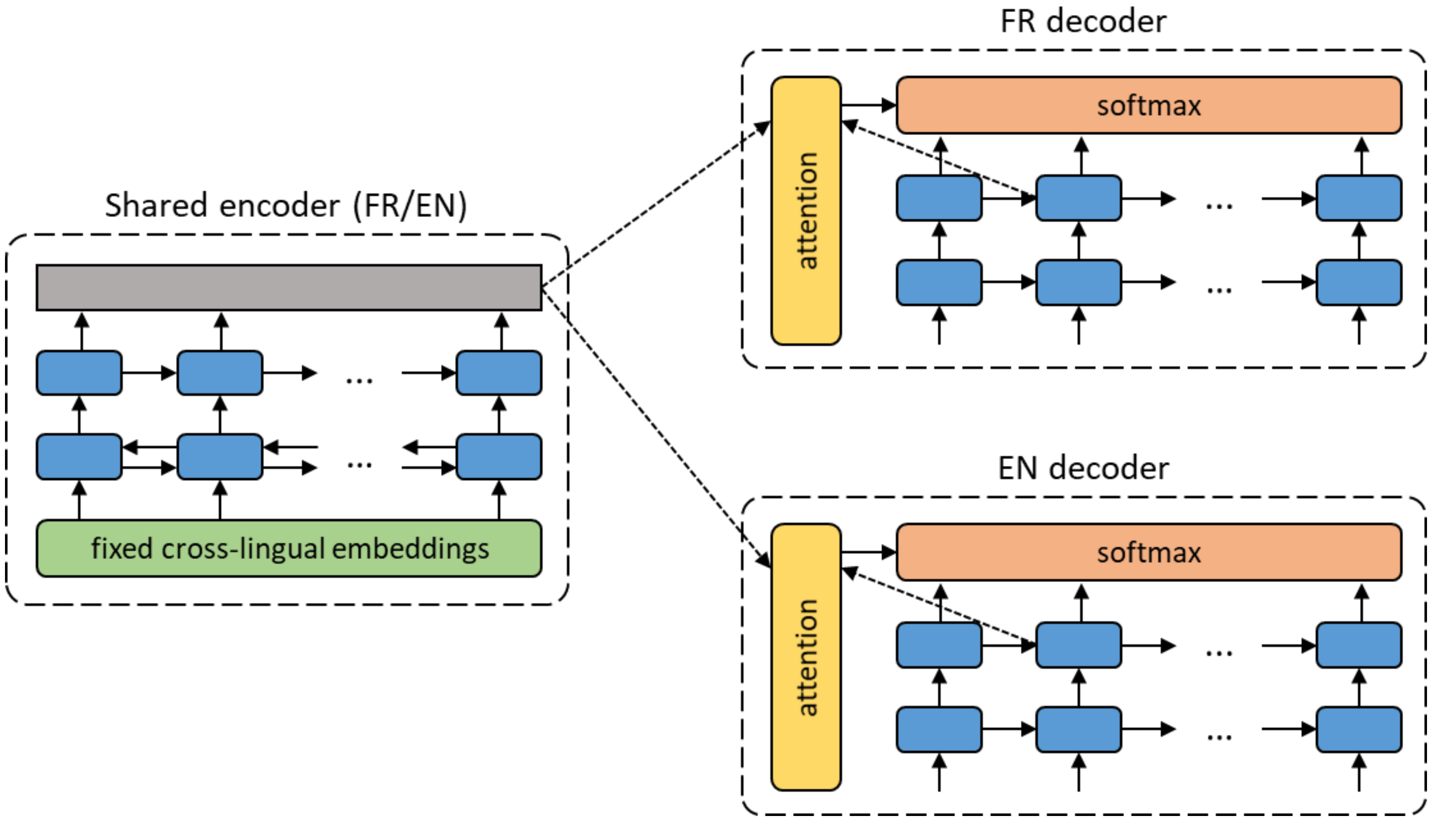
~~*There was a shooting in Los Angeles International Airport.*~~

Unsupervised neural machine translation

Training

- Supervised

Une fusillade a eu lieu à l'aéroport international de Los Angeles.

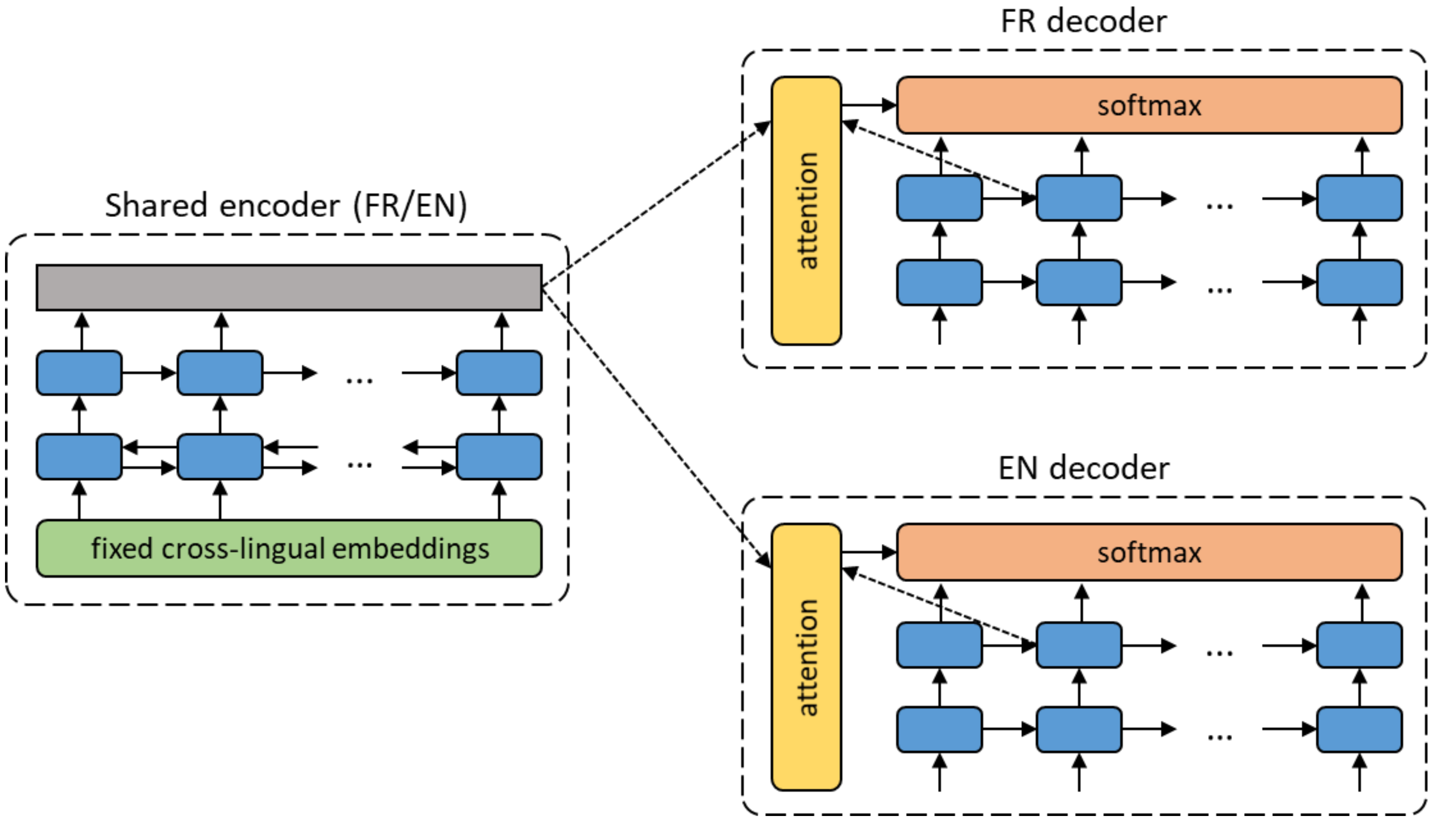


Unsupervised neural machine translation

Training

- ~~Supervised~~

Une fusillade a eu lieu à l'aéroport international de Los Angeles.

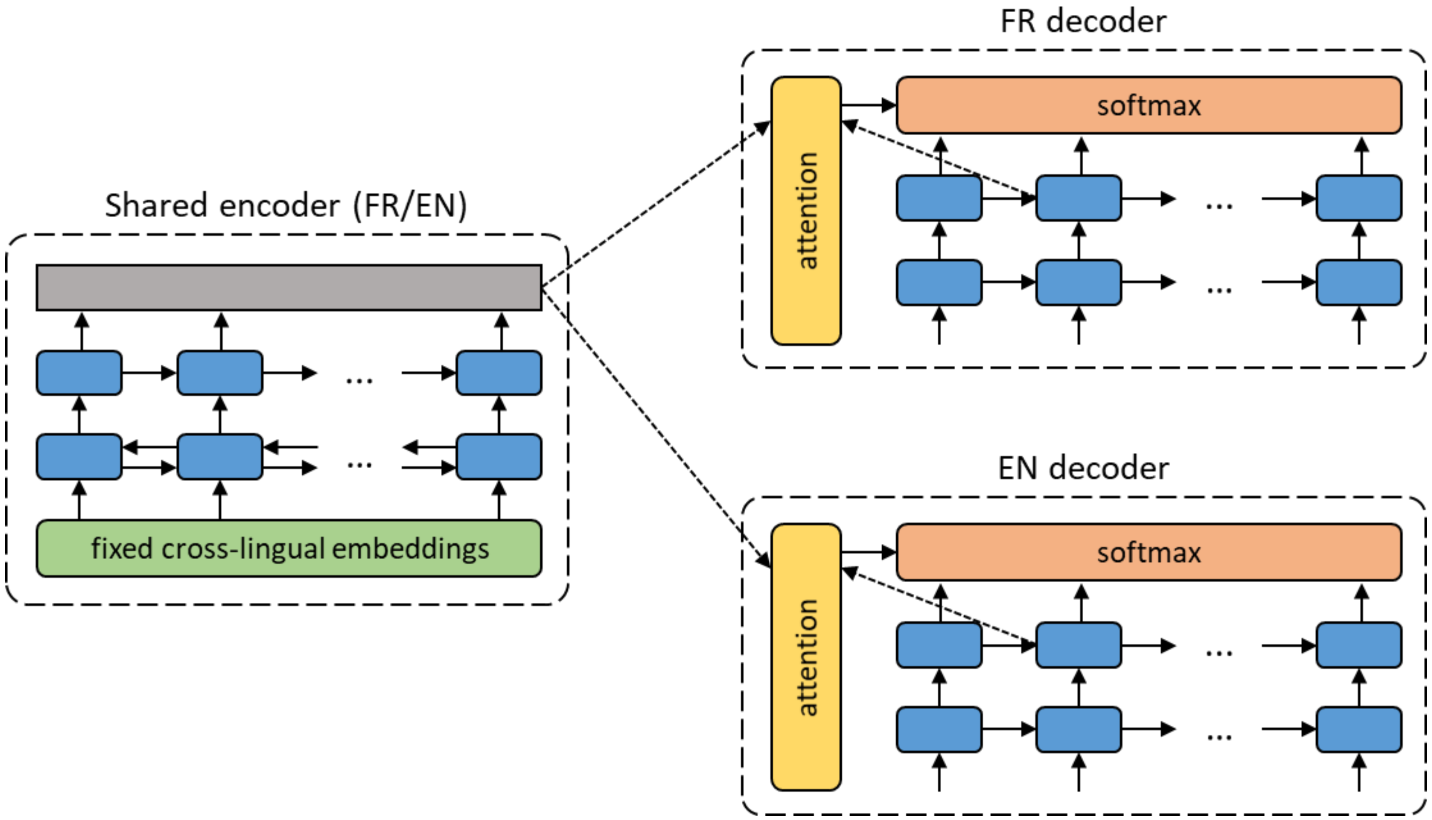


Unsupervised neural machine translation

Training

- ~~Supervised~~

Une fusillade a eu lieu à l'aéroport international de Los Angeles.



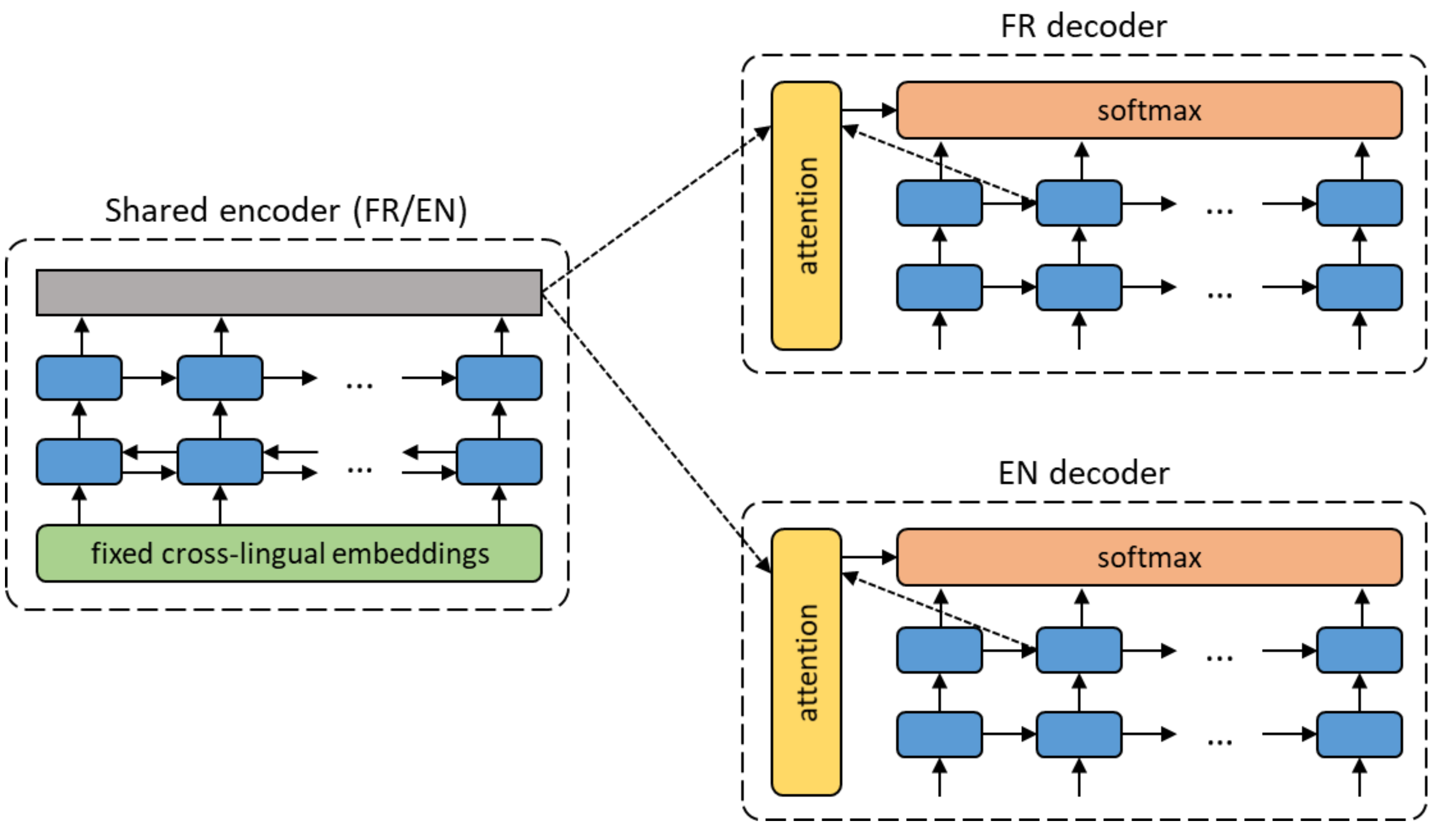
Une fusillade a eu lieu à l'aéroport international de Los Angeles.

Unsupervised neural machine translation

Training

- ~~Supervised~~
- Denoising

Une fusillade a eu lieu à l'aéroport international de Los Angeles.

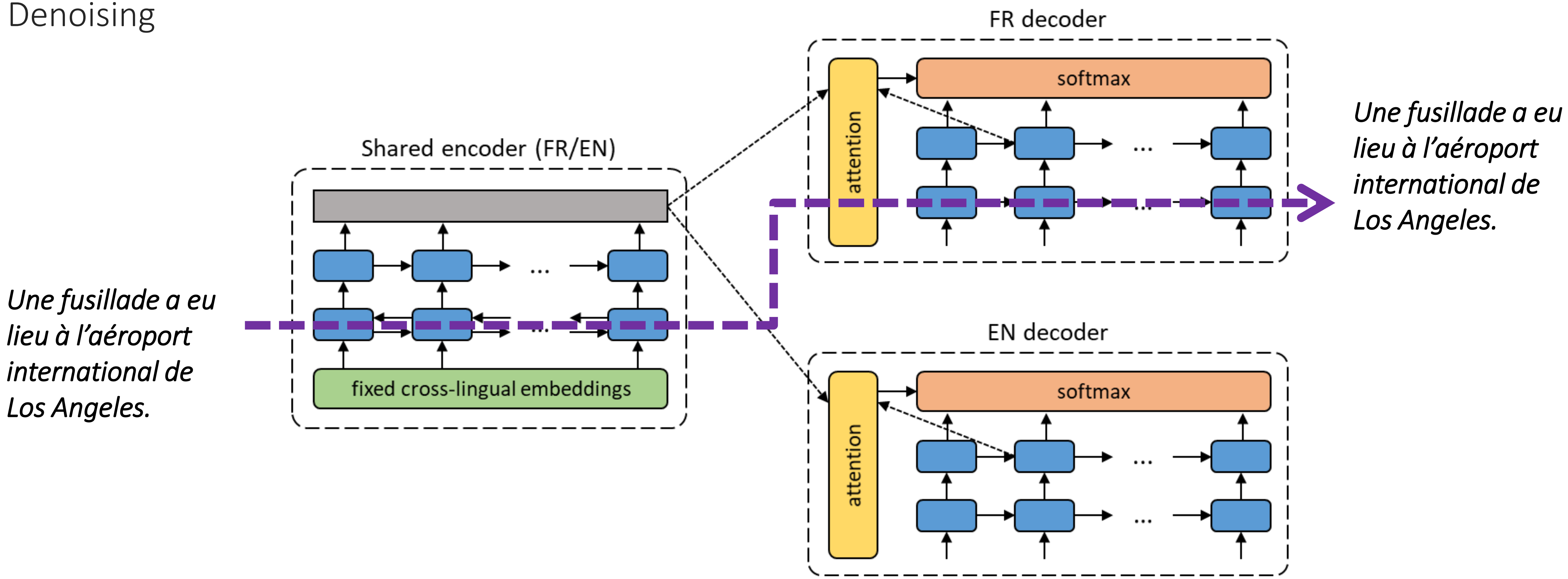


Une fusillade a eu lieu à l'aéroport international de Los Angeles.

Unsupervised neural machine translation

Training

- ~~Supervised~~
- Denoising

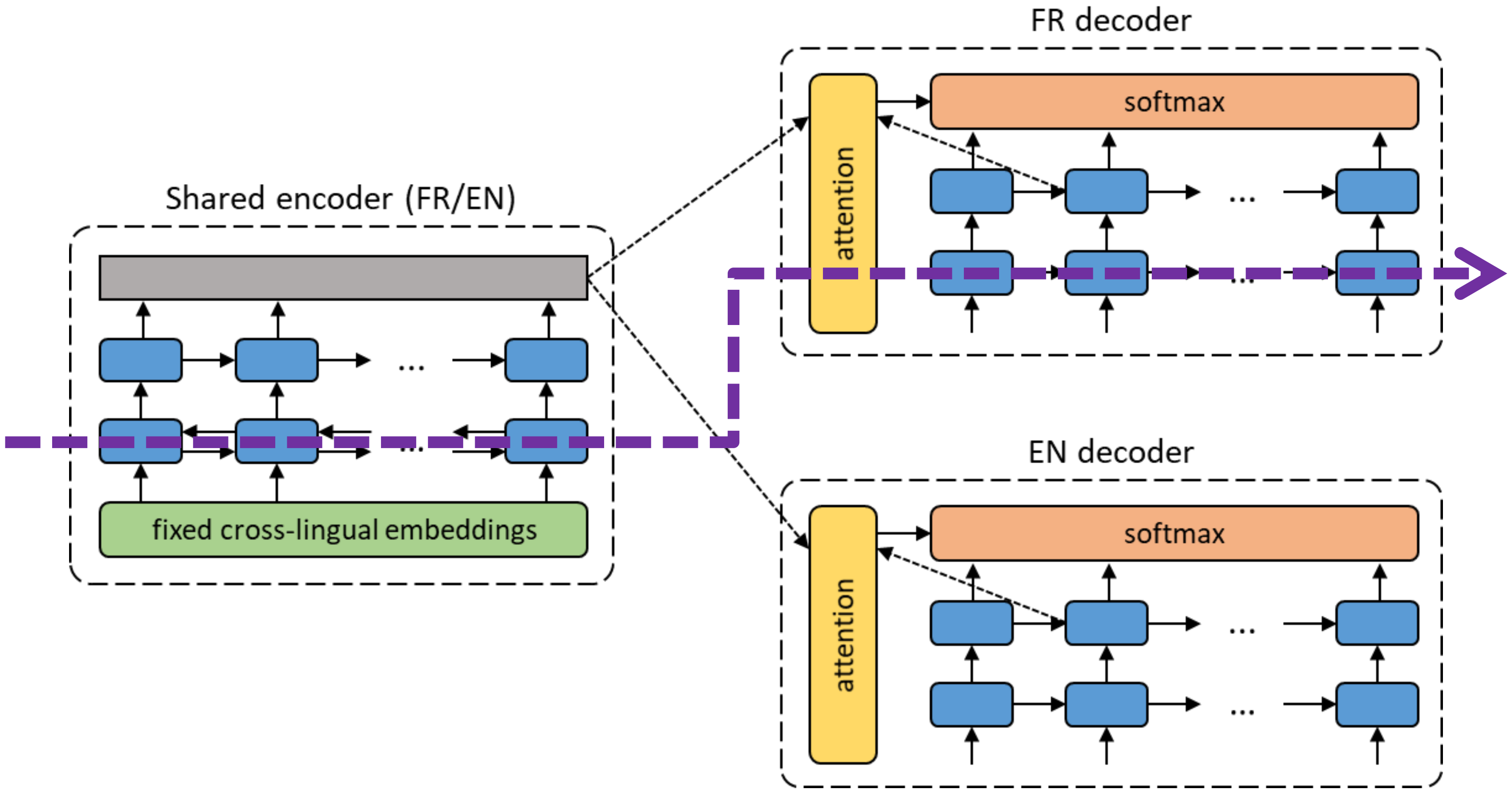


Unsupervised neural machine translation

Training

- ~~Supervised~~
- Denoising

Une lieu fusillade a eu à l'aéroport de Los international Angeles.



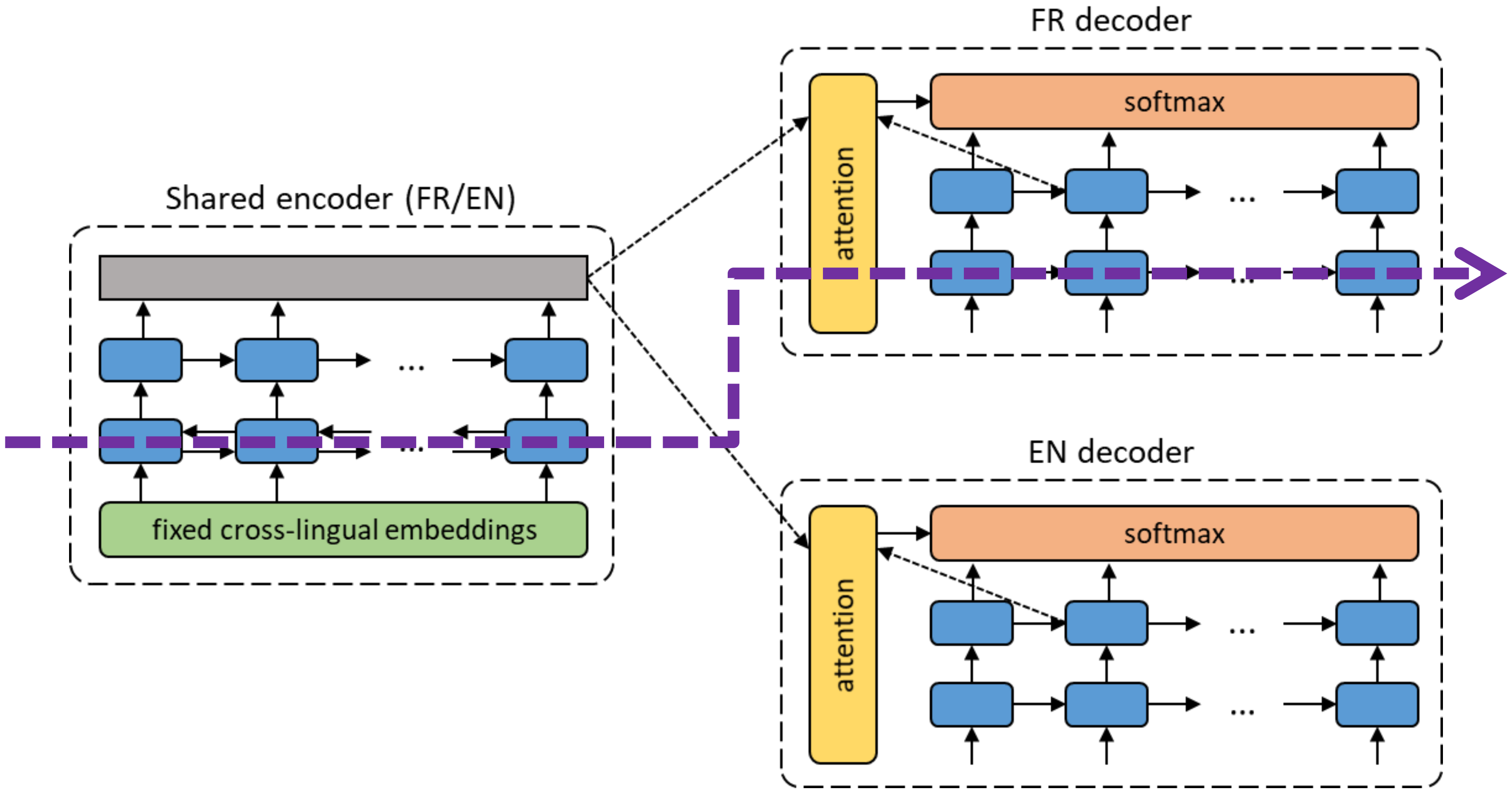
Une fusillade a eu lieu à l'aéroport international de Los Angeles.

Unsupervised neural machine translation

Training

- ~~Supervised~~
- Denoising
- Backtranslation

Une lieu fusillade a eu à l'aéroport de Los international Angeles.



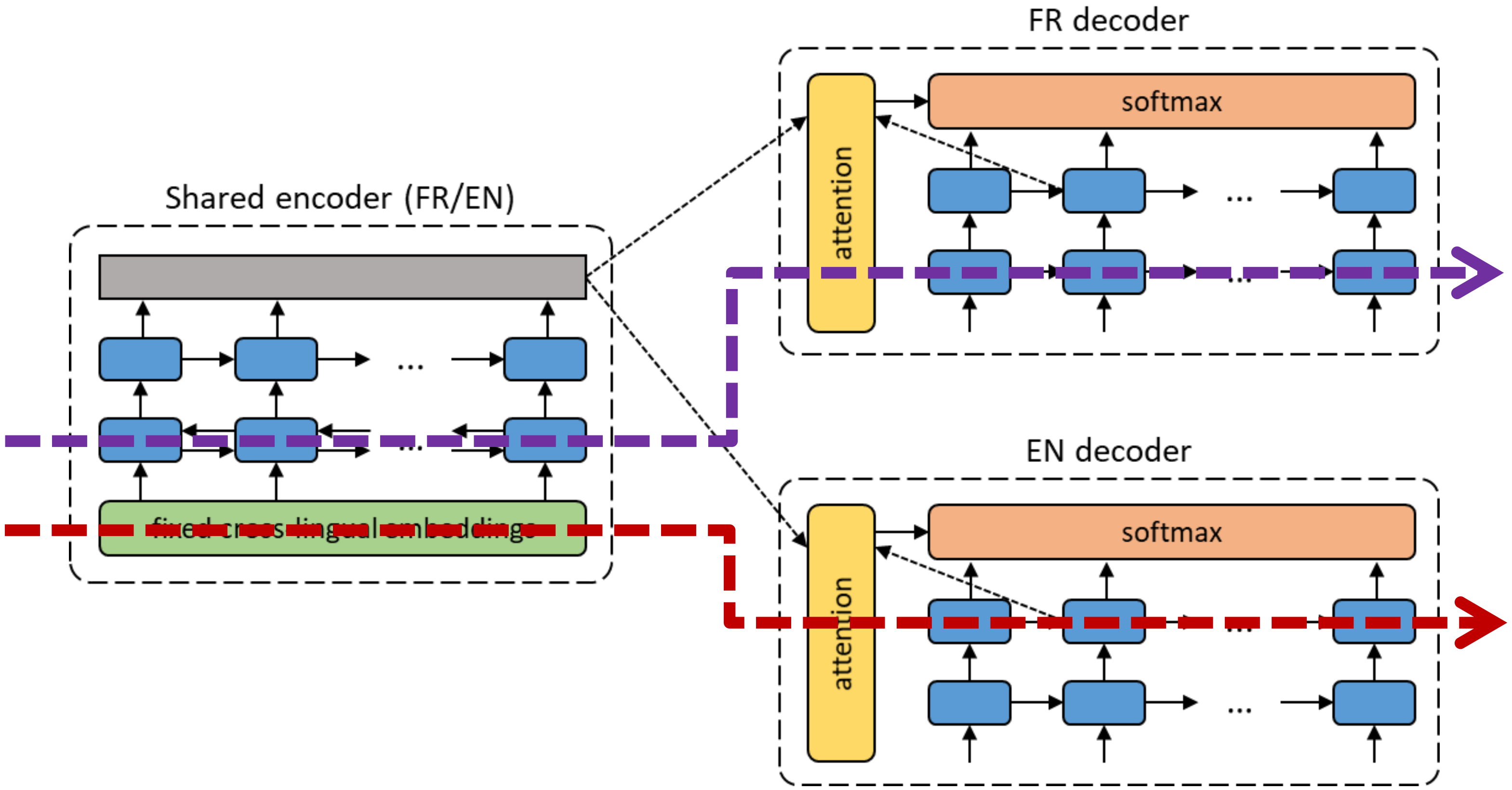
Une fusillade a eu lieu à l'aéroport international de Los Angeles.

Unsupervised neural machine translation

Training

- ~~Supervised~~
- Denoising
- Backtranslation

Une lieu fusillade a eu à l’aéroport de Los international Angeles.

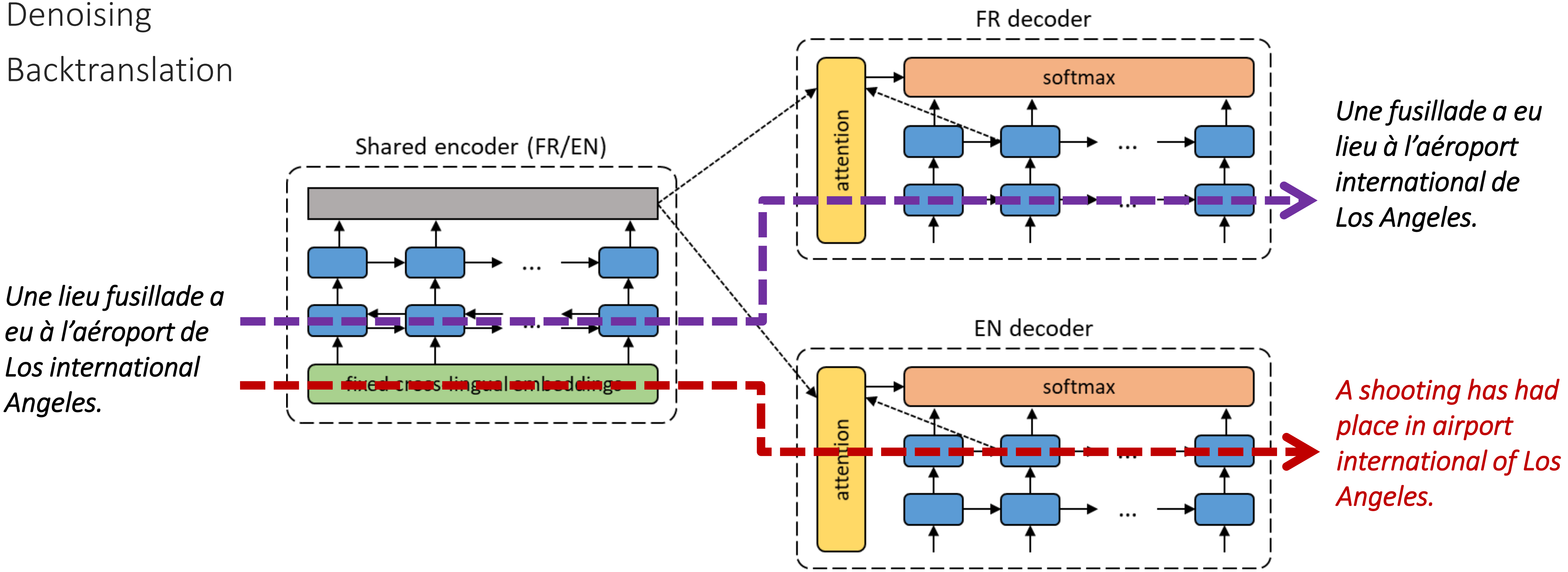


Une fusillade a eu lieu à l’aéroport international de Los Angeles.

Unsupervised neural machine translation

Training

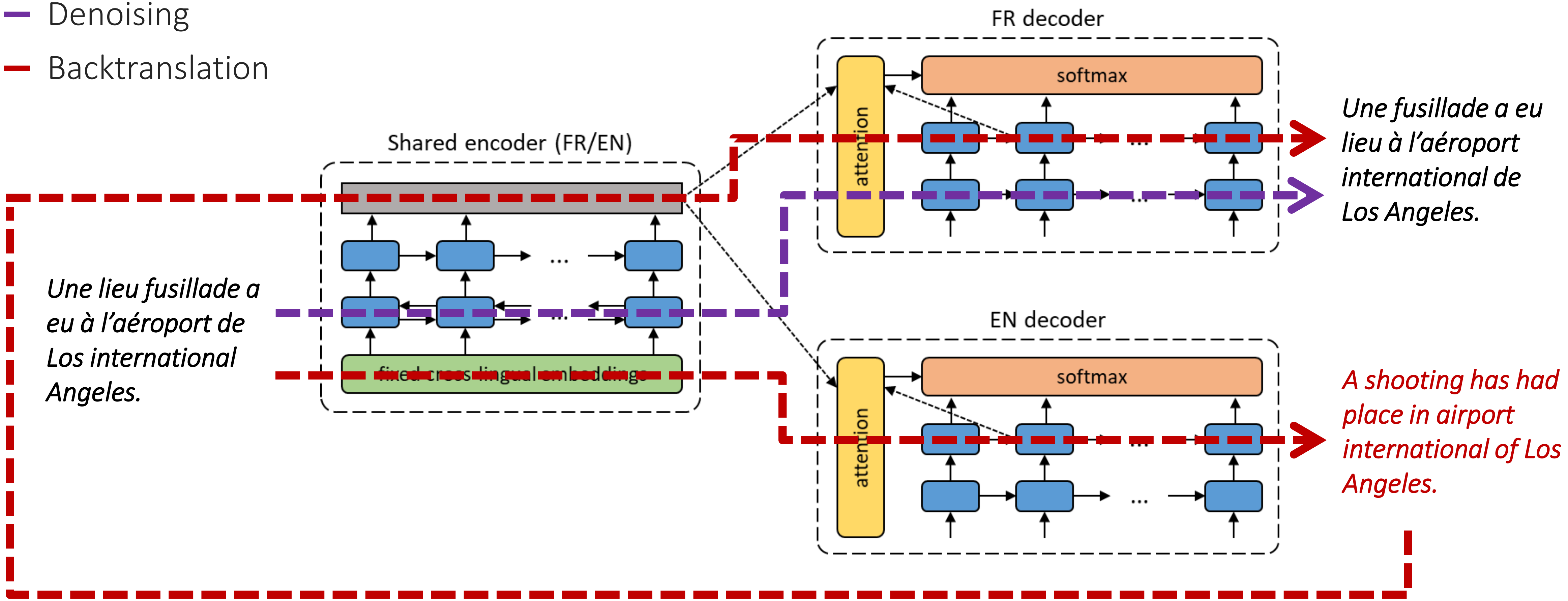
- ~~Supervised~~
- Denoising
- Backtranslation



Unsupervised neural machine translation

Training

- ~~Supervised~~
- Denoising
- Backtranslation



Unsupervised neural machine translation

EXPERIMENTS

Unsupervised neural machine translation

EXPERIMENTS

- Languages: French-English, German-English

Unsupervised neural machine translation

EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl

Unsupervised neural machine translation

EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

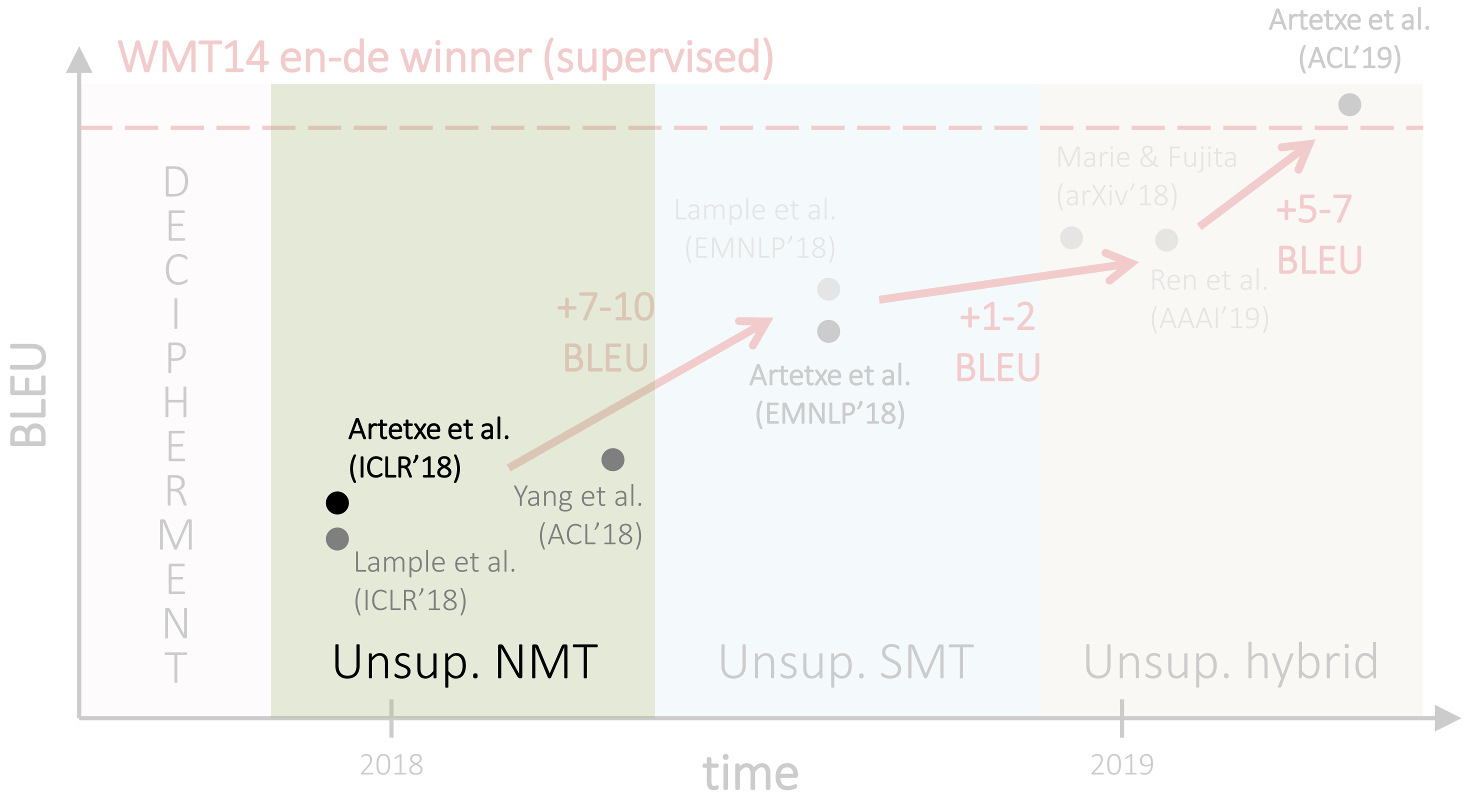
Unsupervised neural machine translation

EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

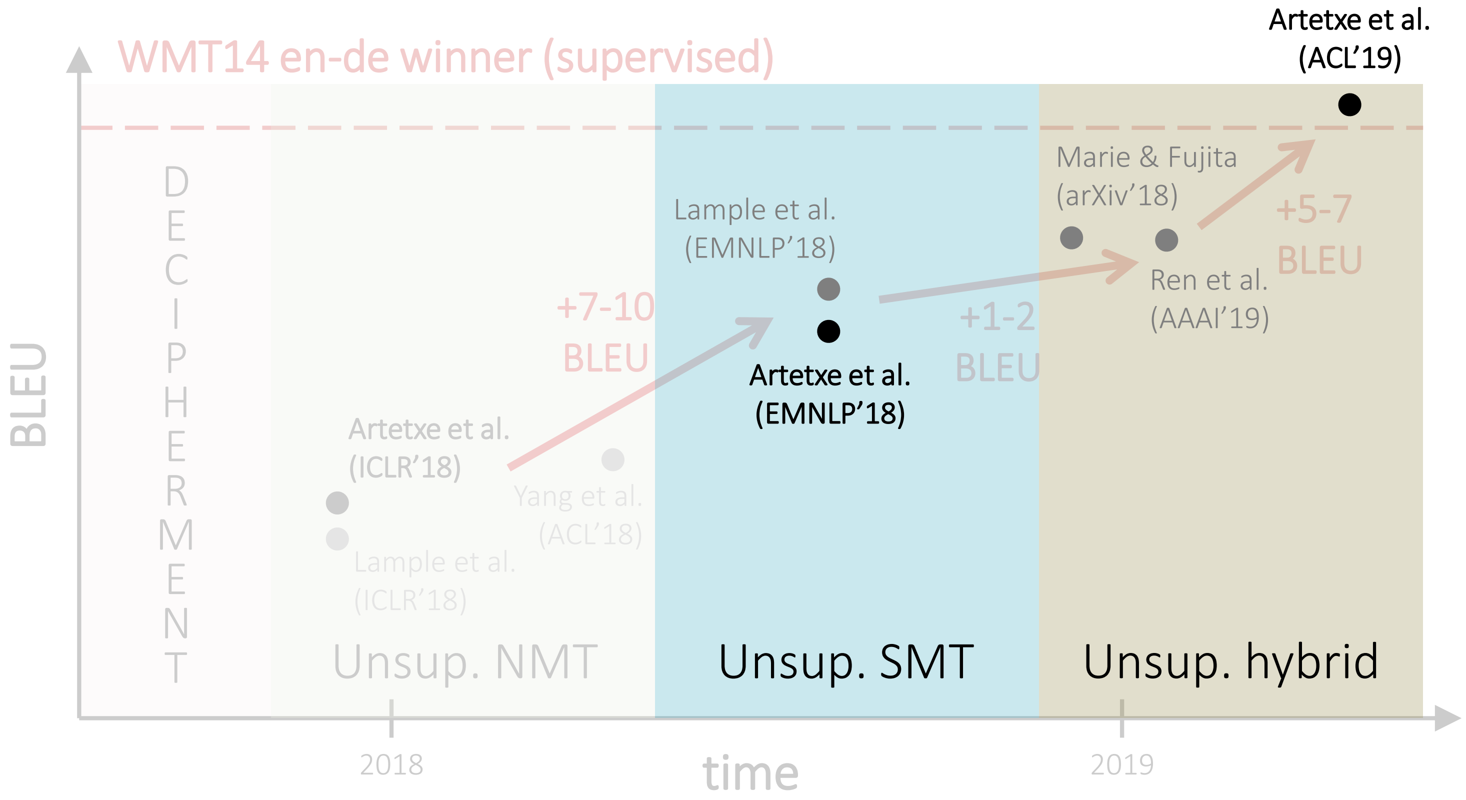
	FR-EN	EN-FR	DE-EN	EN-DE
NMT (ICLR'18)	15.6	15.1	10.2	6.6

Outline

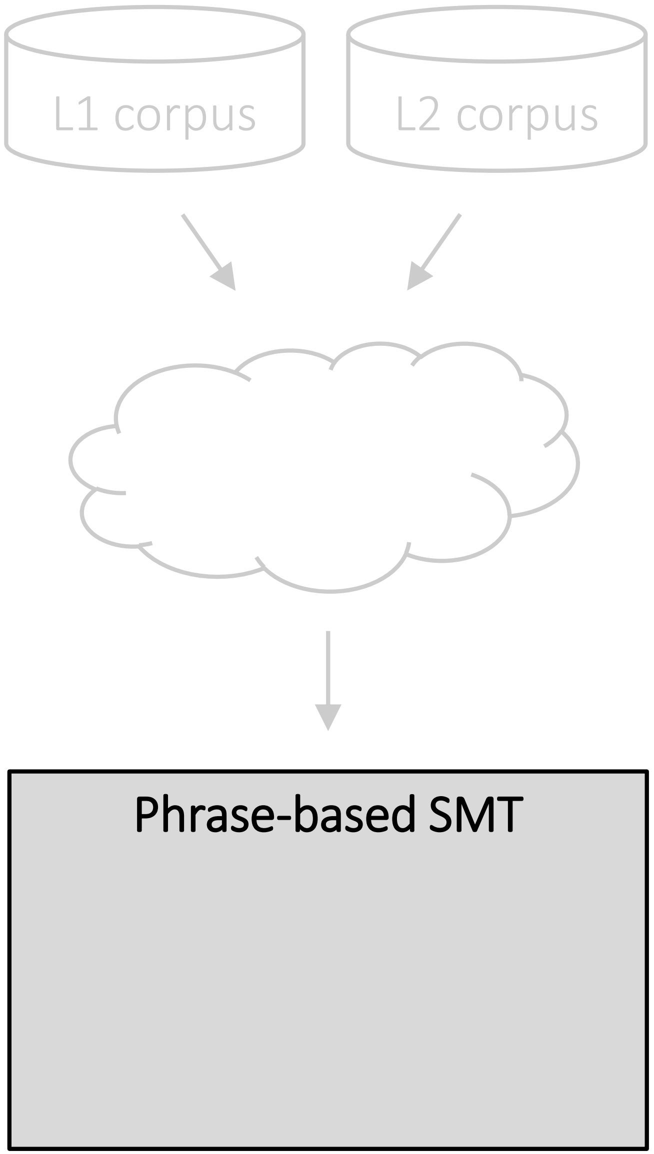


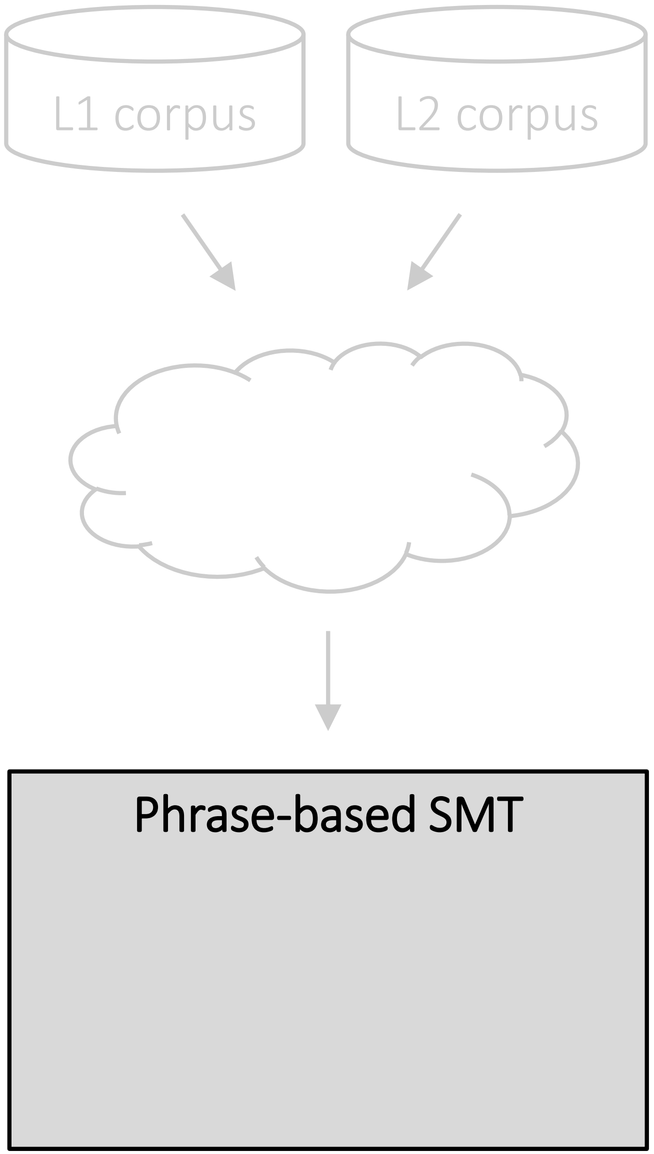
WMT14 en-de winner (supervised)
BLEU
DECIPHERMENT
Artetxe et al.
(ICLR'18)
Lample et al.
(ICLR'18)
Yang et al.
(ACL'18)
+7-10
BLEU
Unsup. NMT
Lample et al.
(EMNLP'18)
Artetxe et al.
(EMNLP'18)
+1-2
BLEU
Unsup. SMT
Marie & Fujita
(arXiv'18)
Ren et al.
(AAAI'19)
+5-7
BLEU
Artetxe et al.
(ACL'19)
Unsup. hybrid
2018
time
2019

Outline

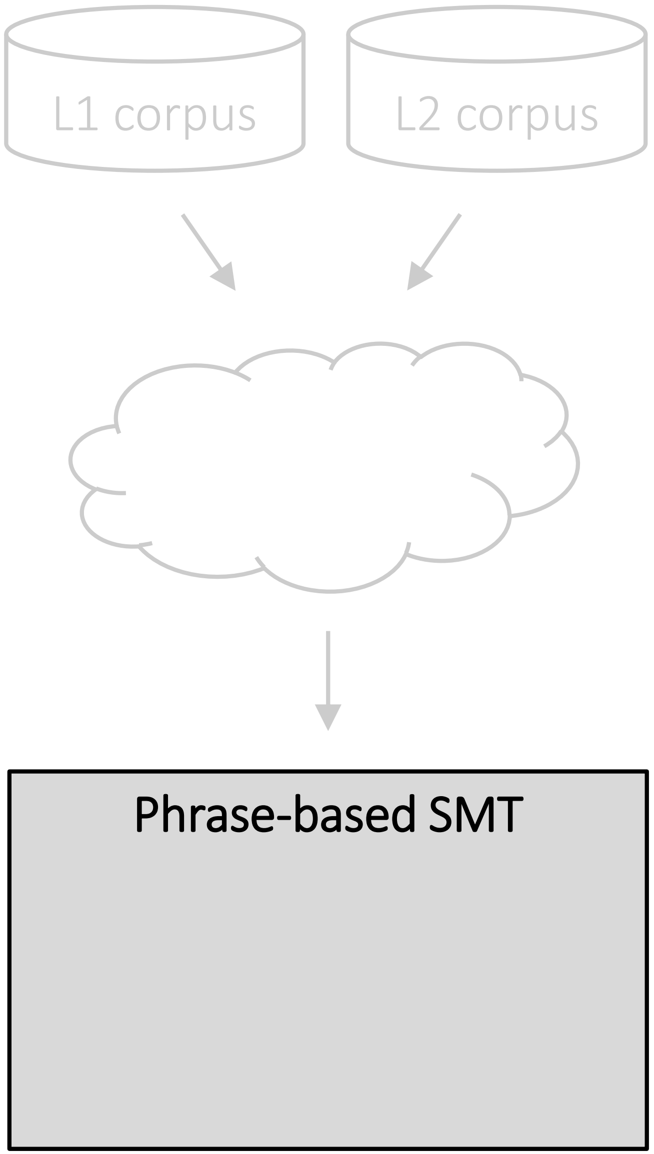


WMT14 en-de winner (supervised)
Artetxe et al.
(ACL'19)
BLEU
DECIPHERMENT
Artetxe et al.
(ICLR'18)
Lampe et al.
(ICLR'18)
Yang et al.
(ACL'18)
+7-10
BLEU
Lample et al.
(EMNLP'18)
Artetxe et al.
(EMNLP'18)
+1-2
BLEU
Marie & Fujita
(arXiv'18)
Ren et al.
(AAAI'19)
+5-7
BLEU
Unsup. NMT
Unsup. SMT
Unsup. hybrid
2018
time
2019


L1 corpus
L2 corpus
Phrase-based SMT

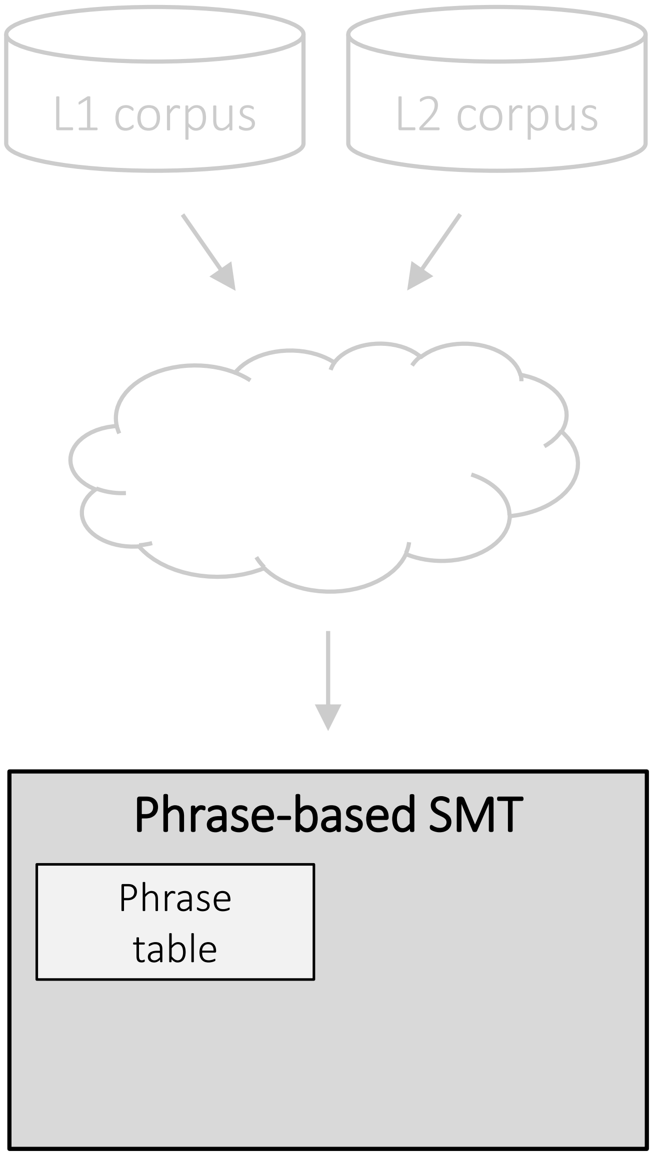


Phrase-based SMT



Phrase-based SMT

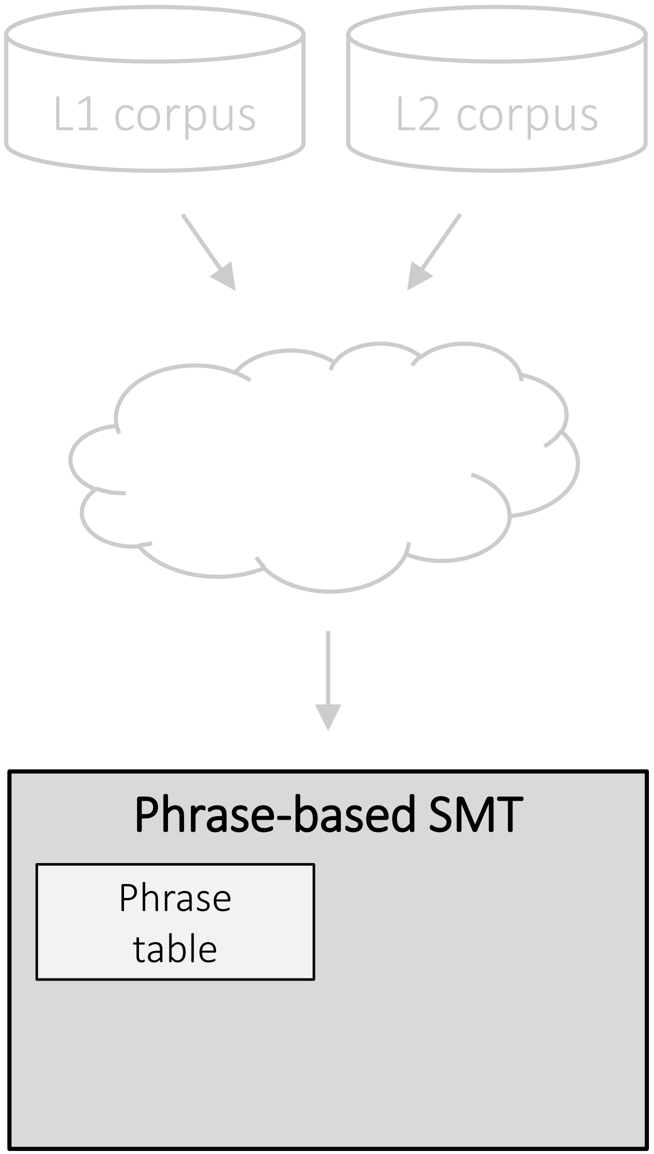
Log-linear model combining



Phrase-based SMT

Log-linear model combining

- Phrase table



Phrase-based SMT

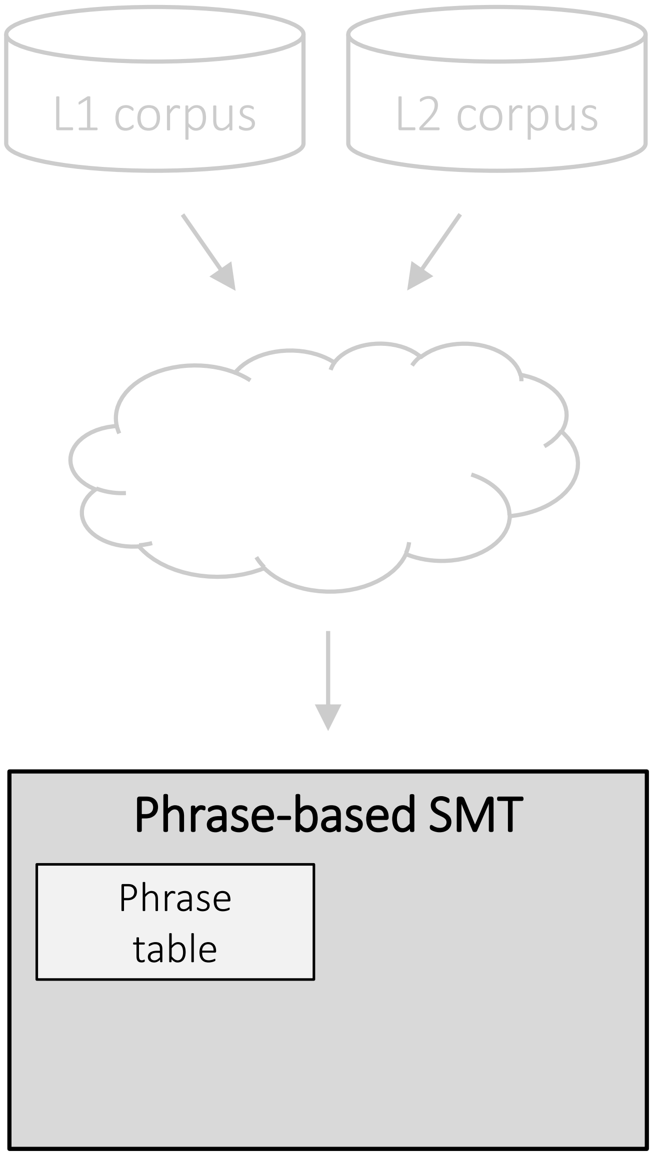
Log-linear model combining

- Phrase table

nire iritziz	in my opinion
nire iritziz	in my view
nire iritziz	I think
opari bat	a present
opari bat	one present
opari bat	a gift

⋮

Phrase-based SMT

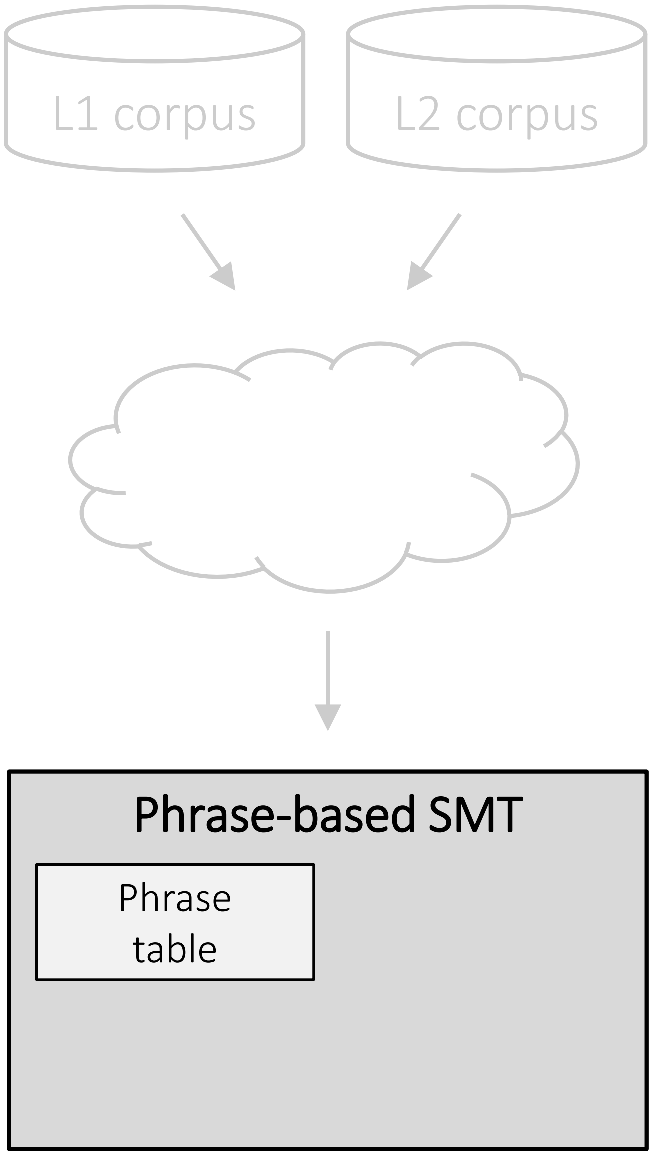


Log-linear model combining

- Phrase table
  - Direct/inverse translation probabilities

		$\phi(\bar{f}\mid\bar{e})$	$\phi(\bar{e}\mid\bar{f})$
nire iritziz	in my opinion	0.54	0.63
nire iritziz	in my view	0.32	0.68
nire iritziz	I think	0.11	0.09
opari bat	a present	0.32	0.56
opari bat	one present	0.14	0.73
opari bat	a gift	0.11	0.49

⋮



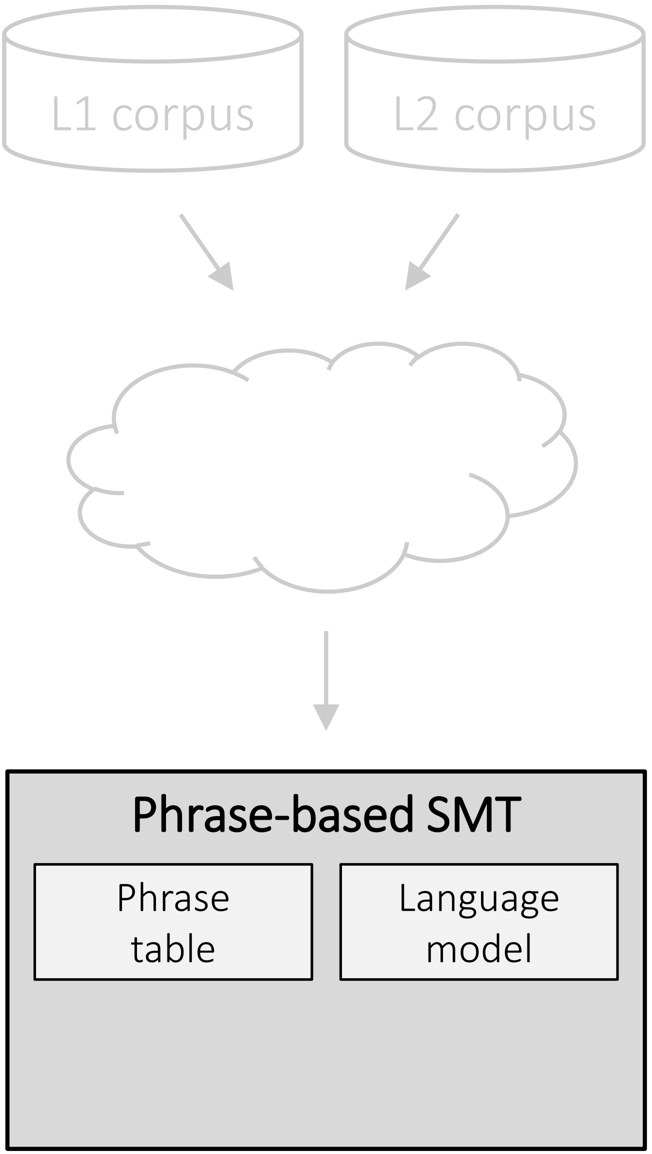
Phrase-based SMT

Log-linear model combining

- Phrase table
    - Direct/inverse translation probabilities
    - Direct/inverse lexical weightings

		$\phi(\bar{f}\|\bar{e})$	$\phi(\bar{e}\|\bar{f})$	$\mathrm{lex}(\bar{f}\|\bar{e})$	$\mathrm{lex}(\bar{e}\|\bar{f})$
nire iritziz	in my opinion	0.54	0.63	0.12	0.15
nire iritziz	in my view	0.32	0.68	0.09	0.16
nire iritziz	I think	0.11	0.09	0.04	0.02
opari bat	a present	0.32	0.56	0.21	0.22
opari bat	one present	0.14	0.73	0.18	0.32
opari bat	a gift	0.11	0.49	0.11	0.13

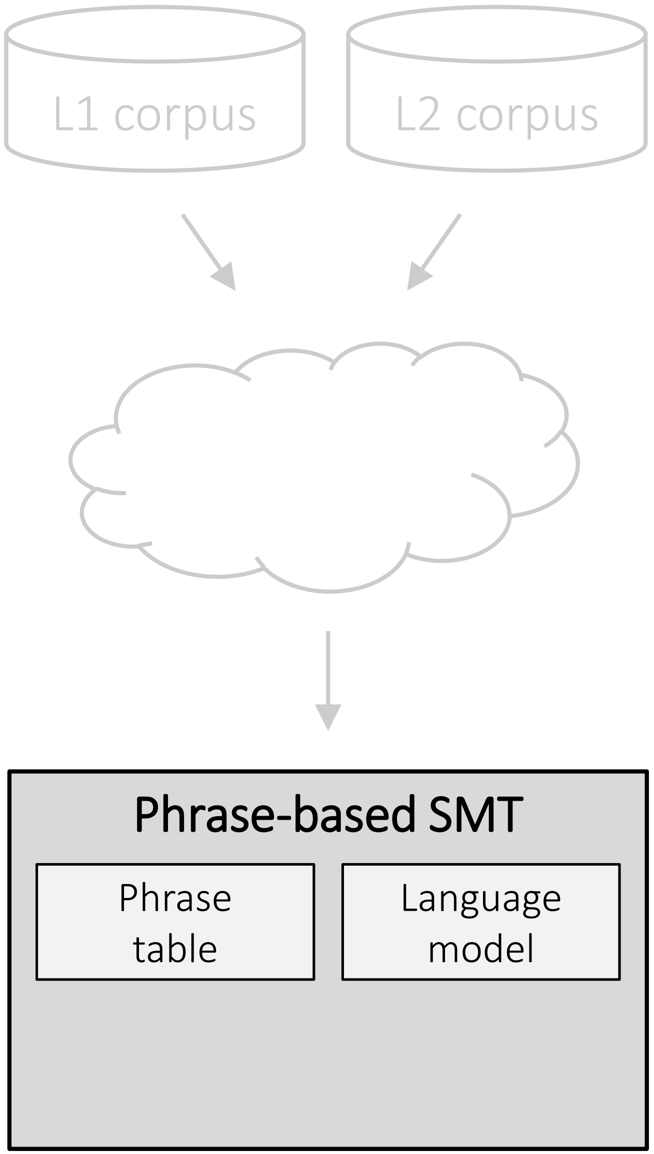
⋮



Phrase-based SMT

Log-linear model combining

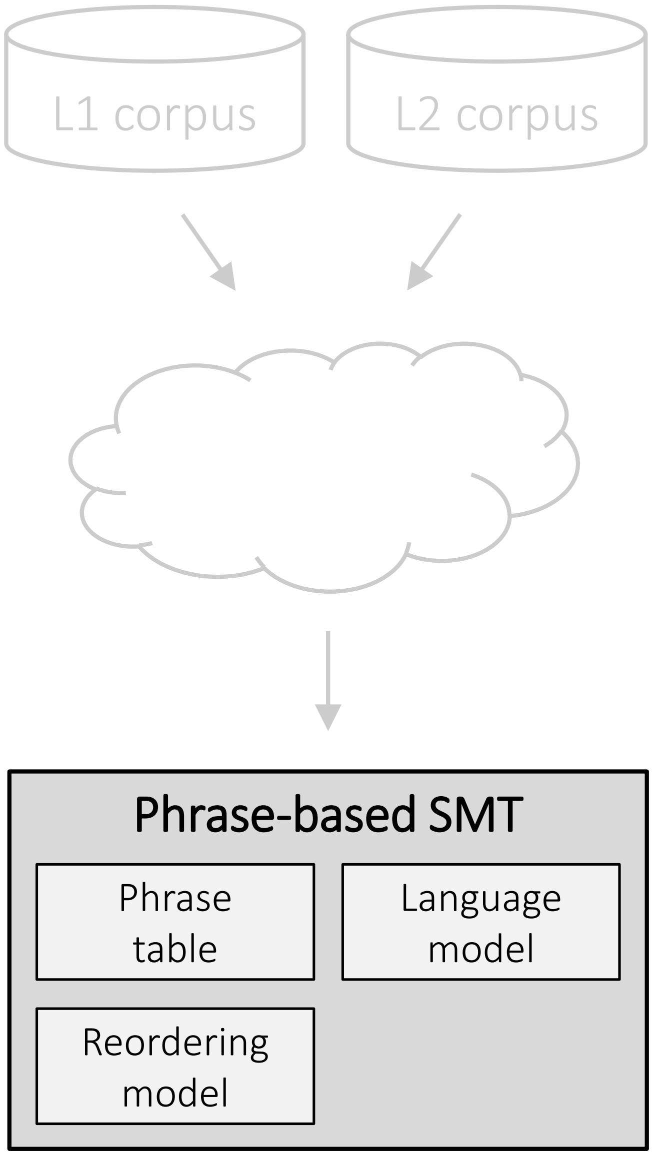
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model



Phrase-based SMT

Log-linear model combining

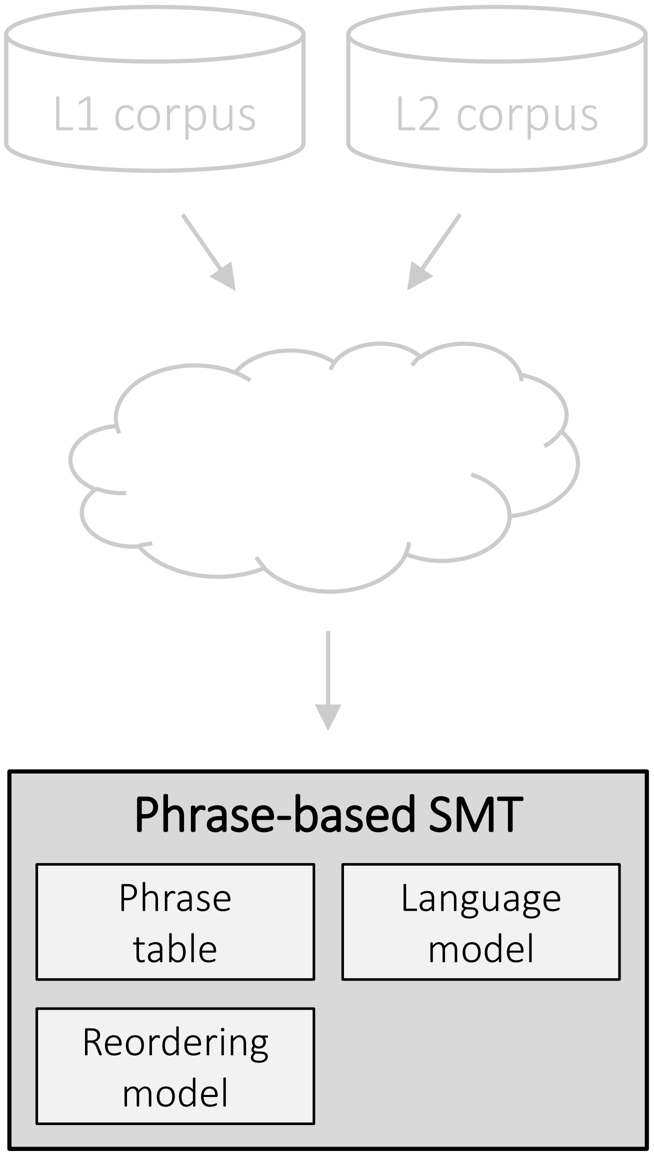
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing



Phrase-based SMT

Log-linear model combining

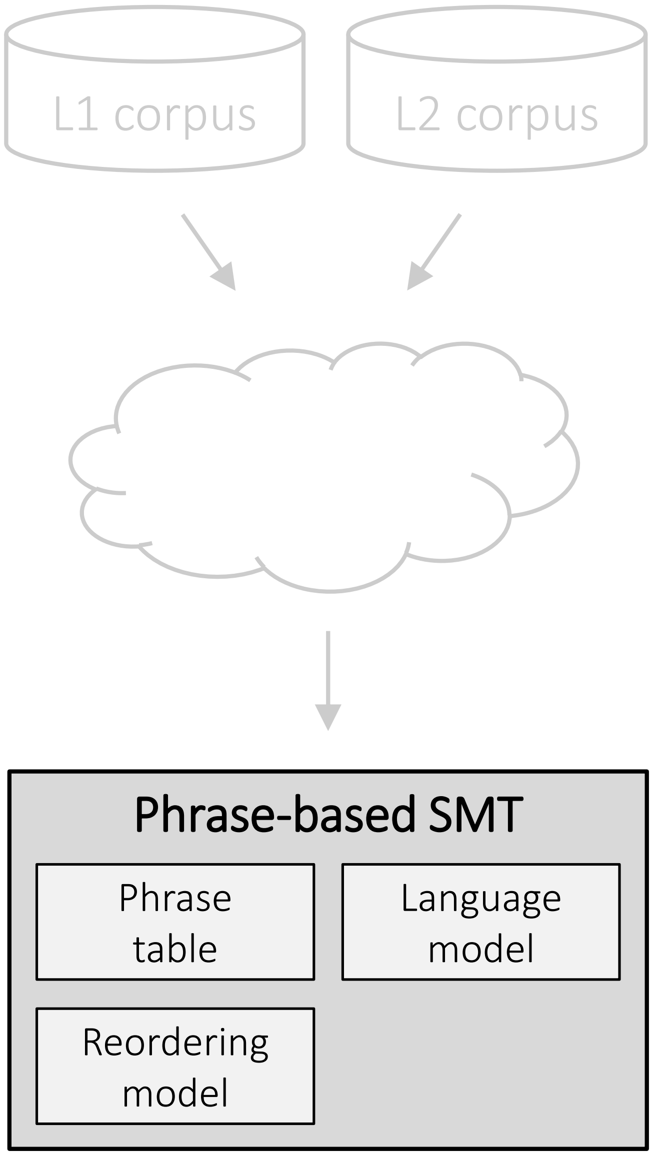
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model



Phrase-based SMT

Log-linear model combining

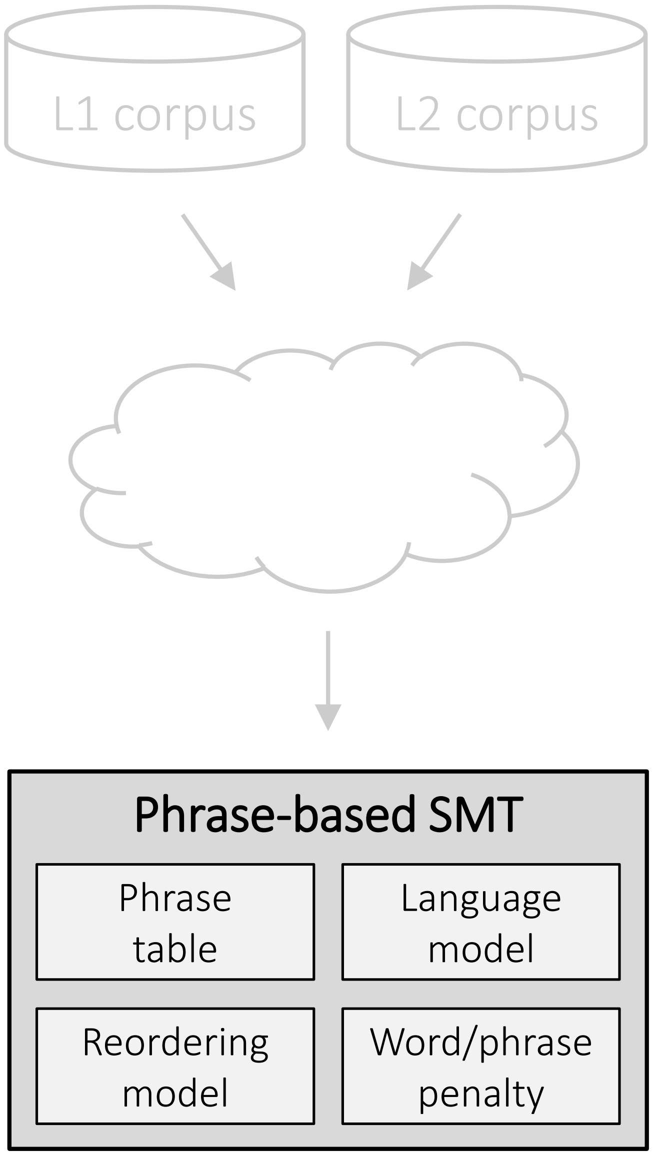
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)



Phrase-based SMT

Log-linear model combining

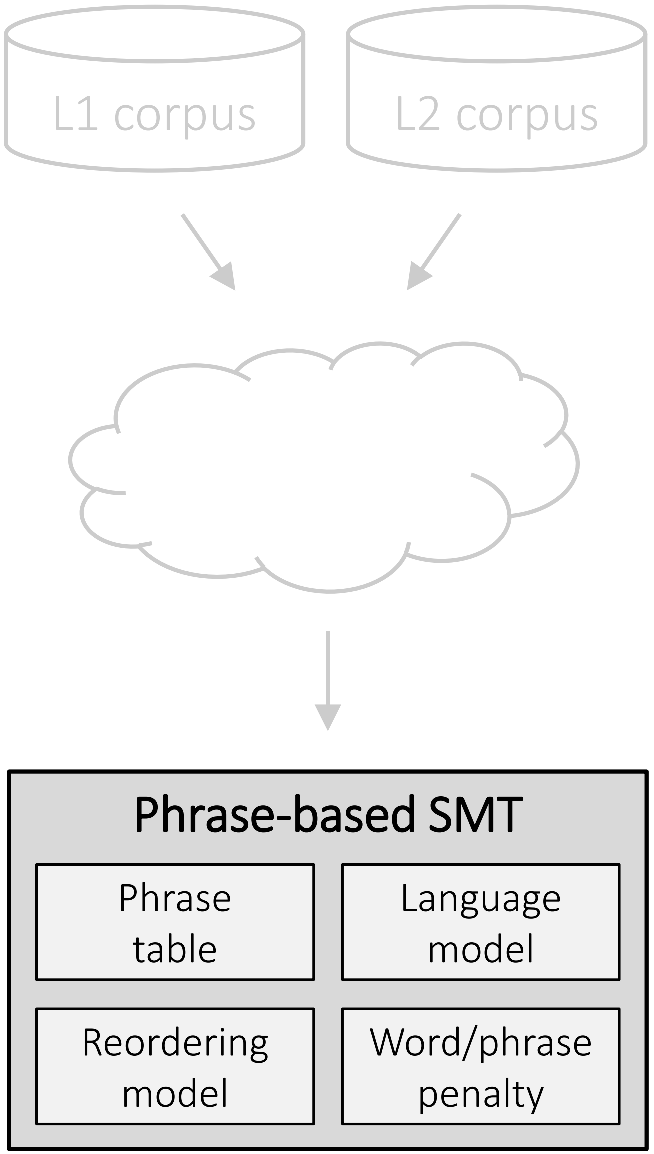
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - Lexical reordering model



Phrase-based SMT

Log-linear model combining

- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - Lexical reordering model
- Word/phrase penalty

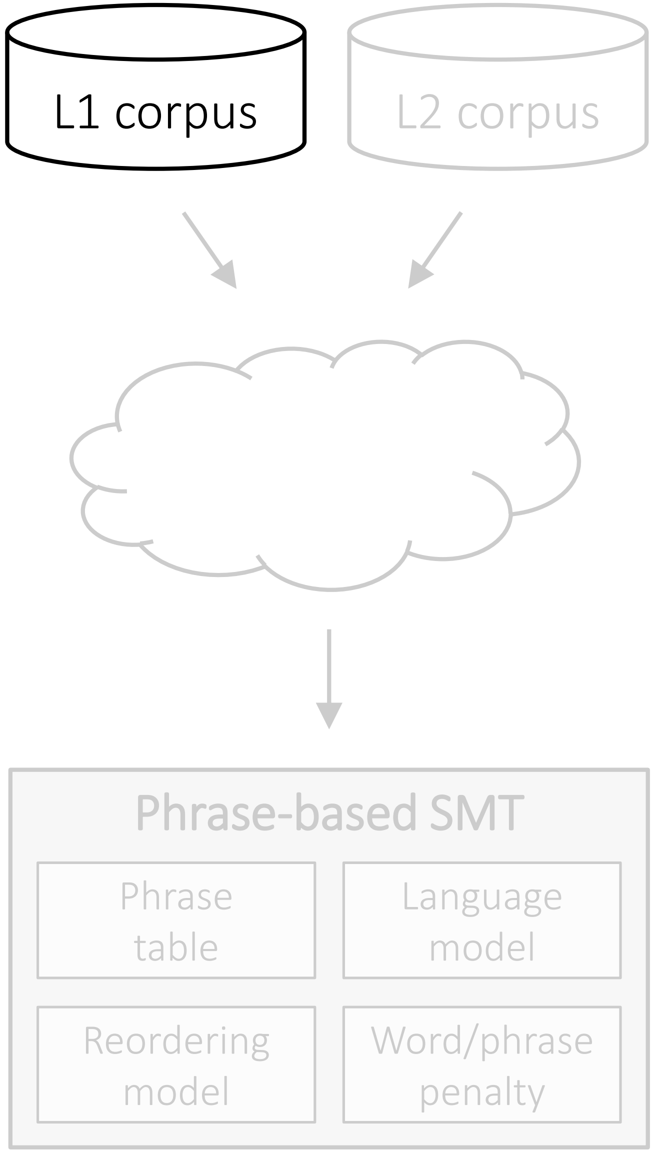


Phrase-based SMT

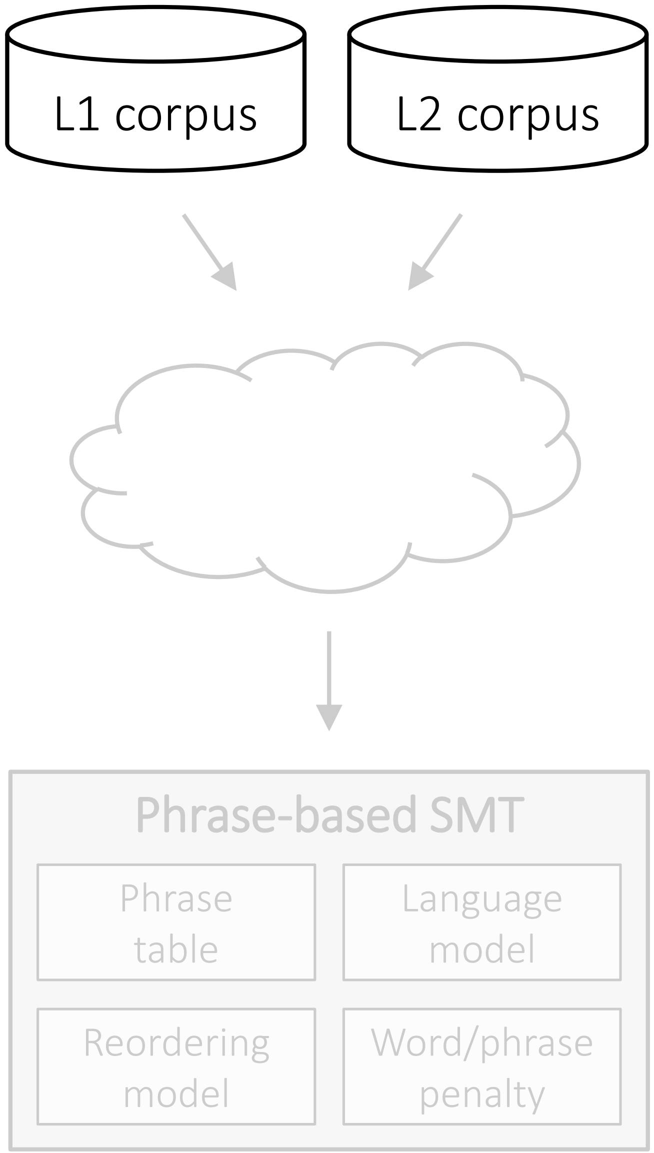
Log-linear model combining

- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - Lexical reordering model
- Word/phrase penalty
  - Fixed score to control the length of the output

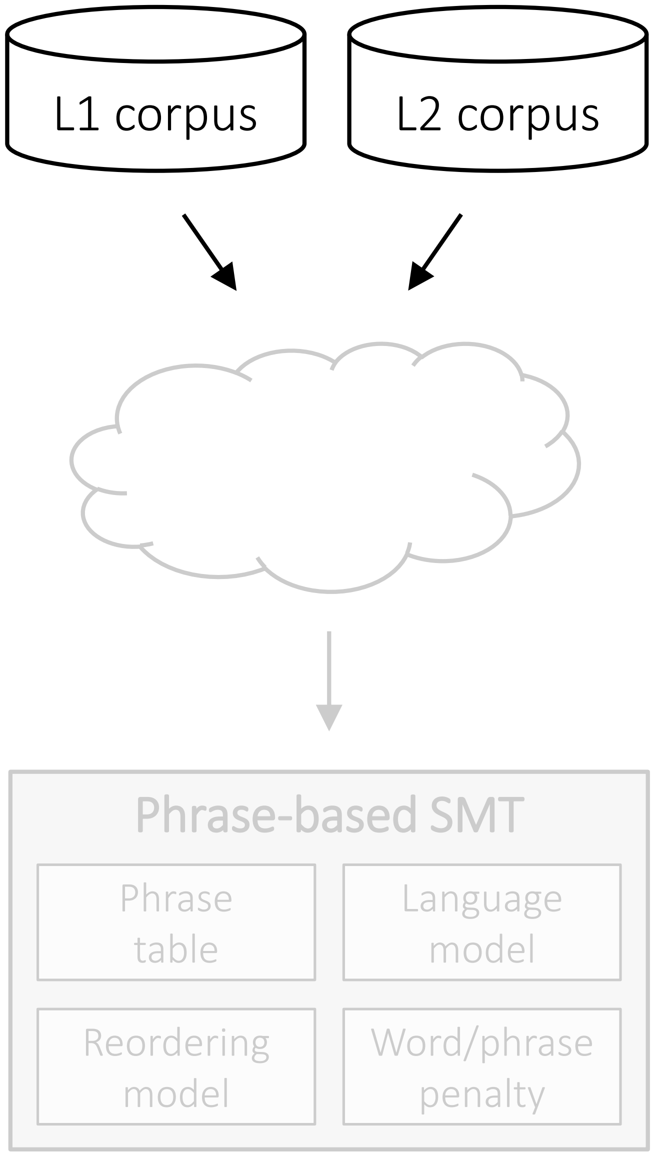
- Phrase table
    - Direct/inverse translation probabilities
    - Direct/inverse lexical weightings
- Language model
    - N-gram frequency counts with back-off and smoothing
- Reordering model
    - Distortion model (distance based)
    - Lexical reordering model
- Word/phrase penalty
    - Fixed score to control the length of the output



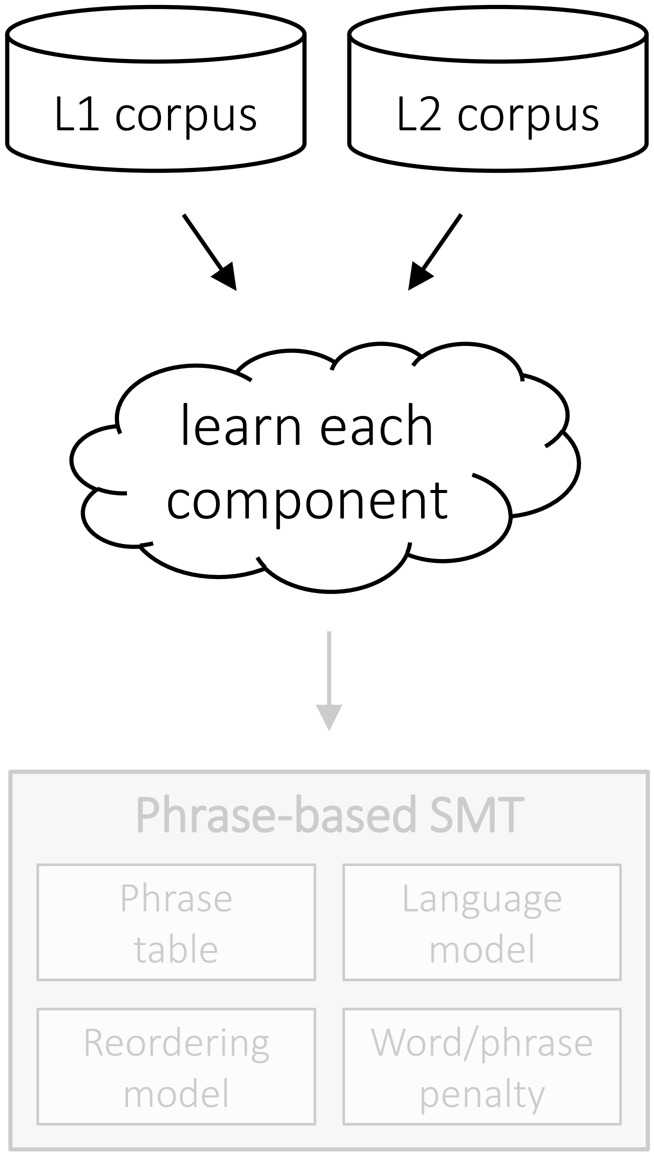
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - Lexical reordering model
- Word/phrase penalty
  - Fixed score to control the length of the output



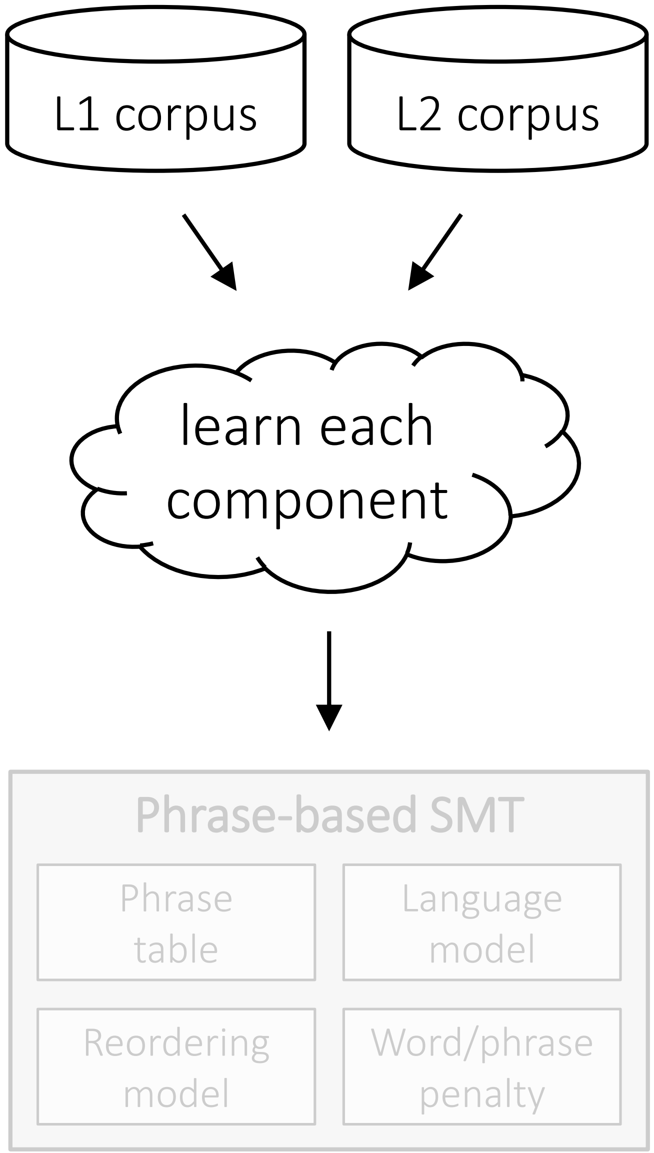
- Phrase table
    - Direct/inverse translation probabilities
    - Direct/inverse lexical weightings
- Language model
    - N-gram frequency counts with back-off and smoothing
- Reordering model
    - Distortion model (distance based)
    - Lexical reordering model
- Word/phrase penalty
    - Fixed score to control the length of the output



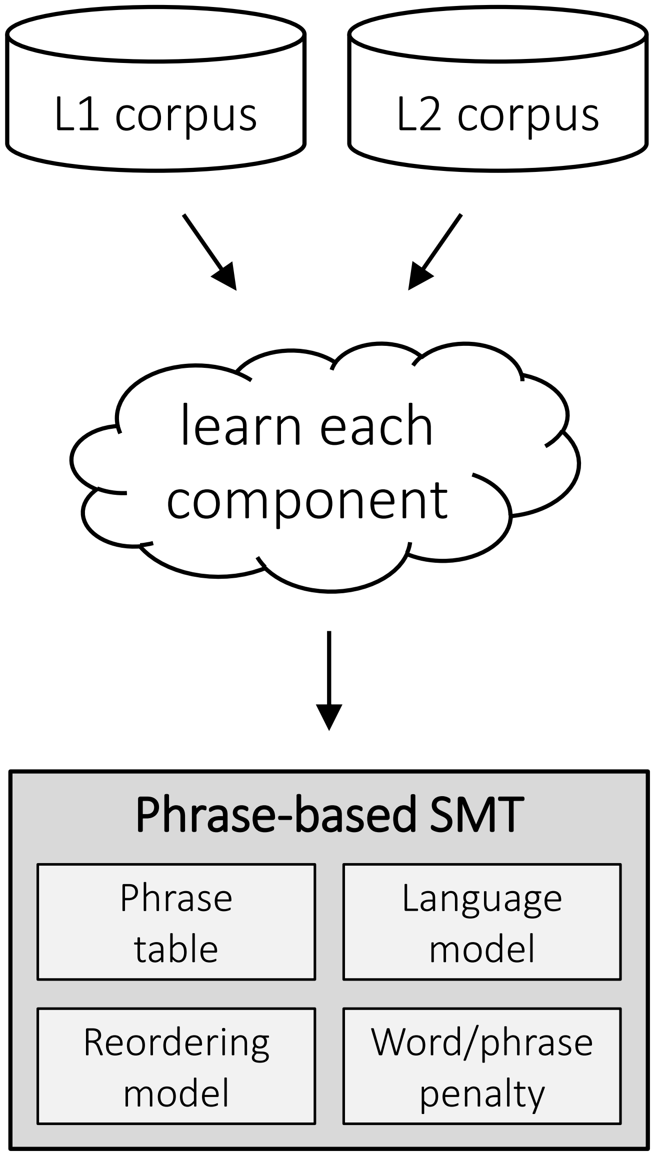
- Phrase table
    - Direct/inverse translation probabilities
    - Direct/inverse lexical weightings
- Language model
    - N-gram frequency counts with back-off and smoothing
- Reordering model
    - Distortion model (distance based)
    - Lexical reordering model
- Word/phrase penalty
    - Fixed score to control the length of the output



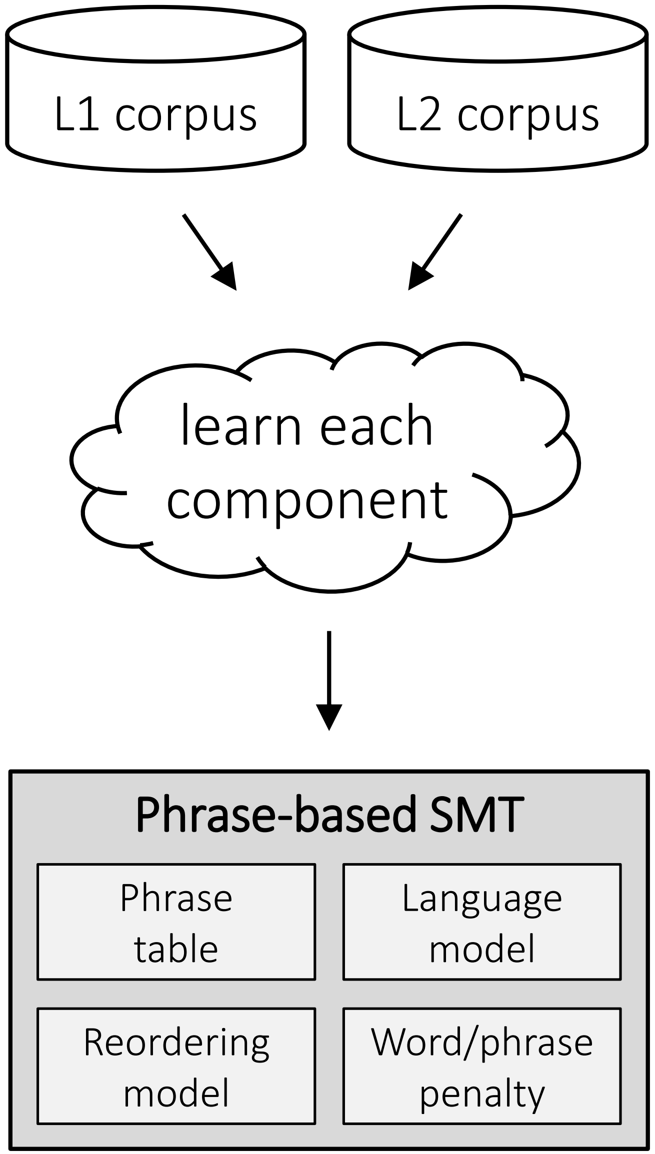
- Phrase table
    - Direct/inverse translation probabilities
    - Direct/inverse lexical weightings
- Language model
    - N-gram frequency counts with back-off and smoothing
- Reordering model
    - Distortion model (distance based)
    - Lexical reordering model
- Word/phrase penalty
    - Fixed score to control the length of the output



- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - Lexical reordering model
- Word/phrase penalty
  - Fixed score to control the length of the output

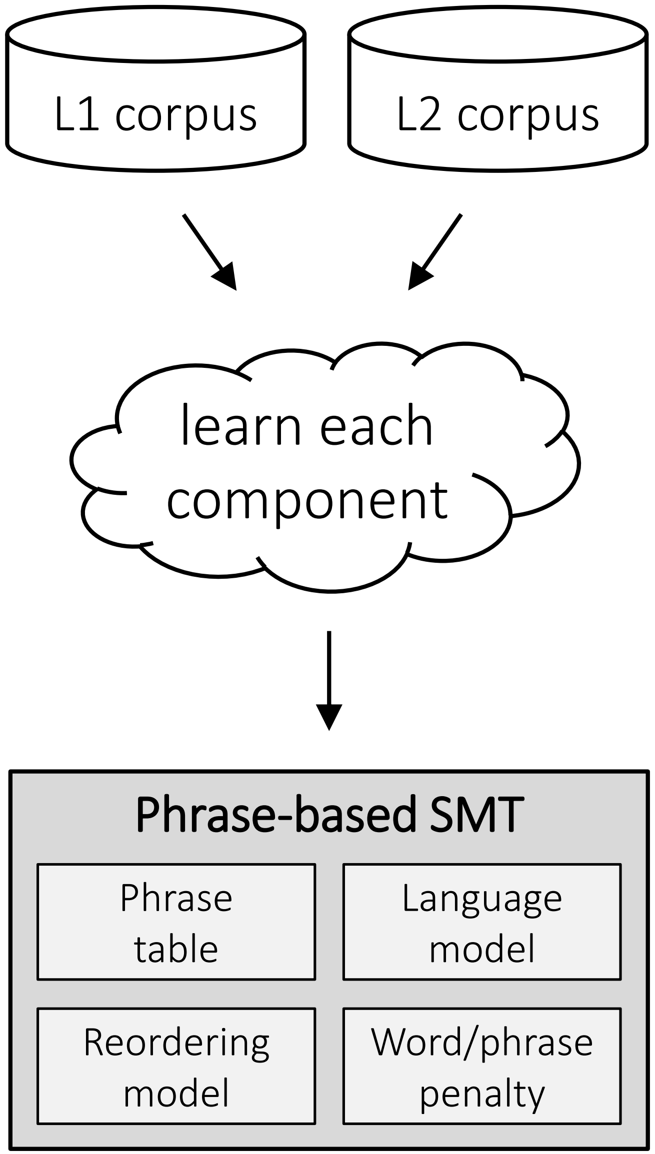


- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - Lexical reordering model
- Word/phrase penalty
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

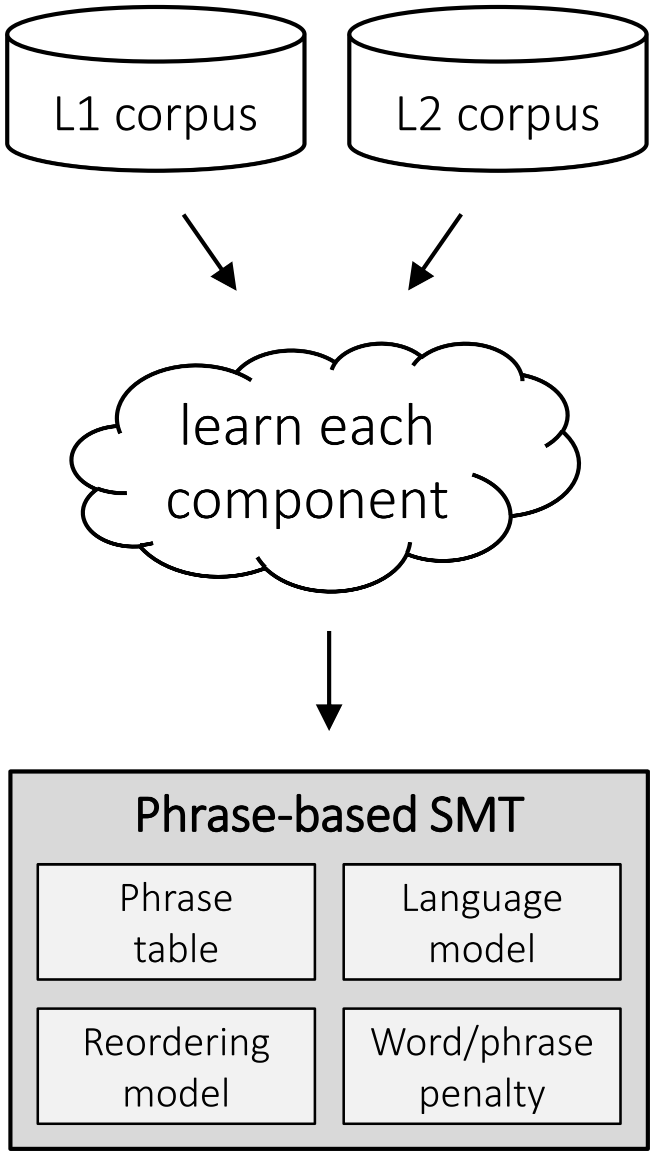
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - Lexical reordering model
- Word/phrase penalty
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - Lexical reordering model
- Word/phrase penalty
  - Fixed score to control the length of the output

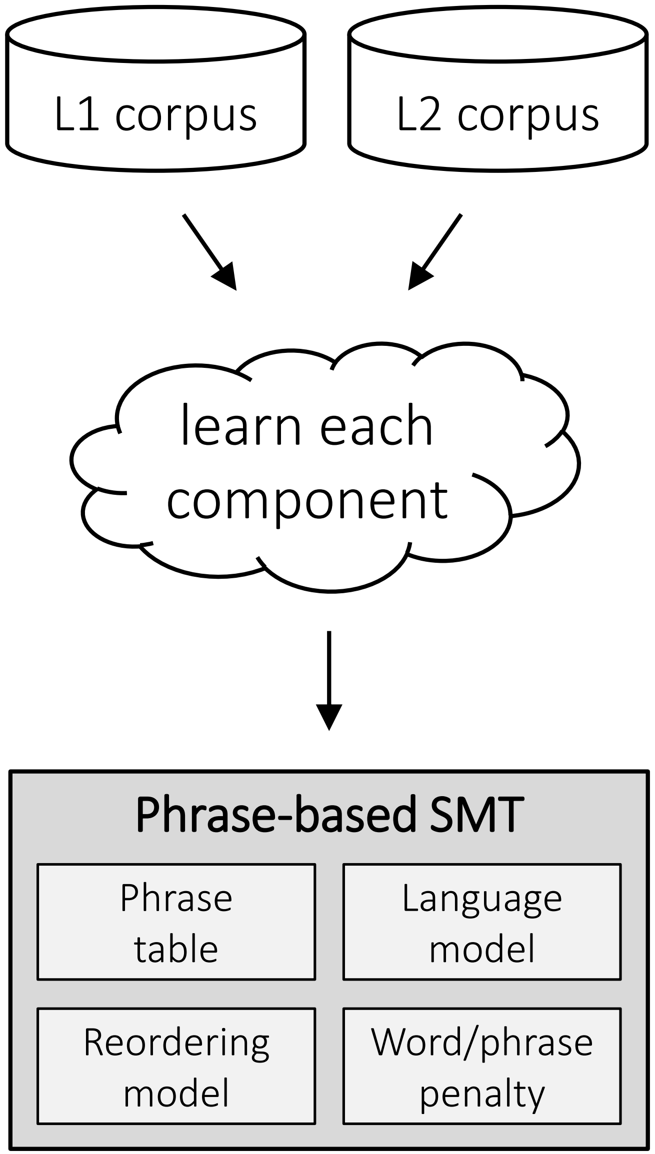


Unsupervised phrase-based SMT

Learn components from monolingual corpora

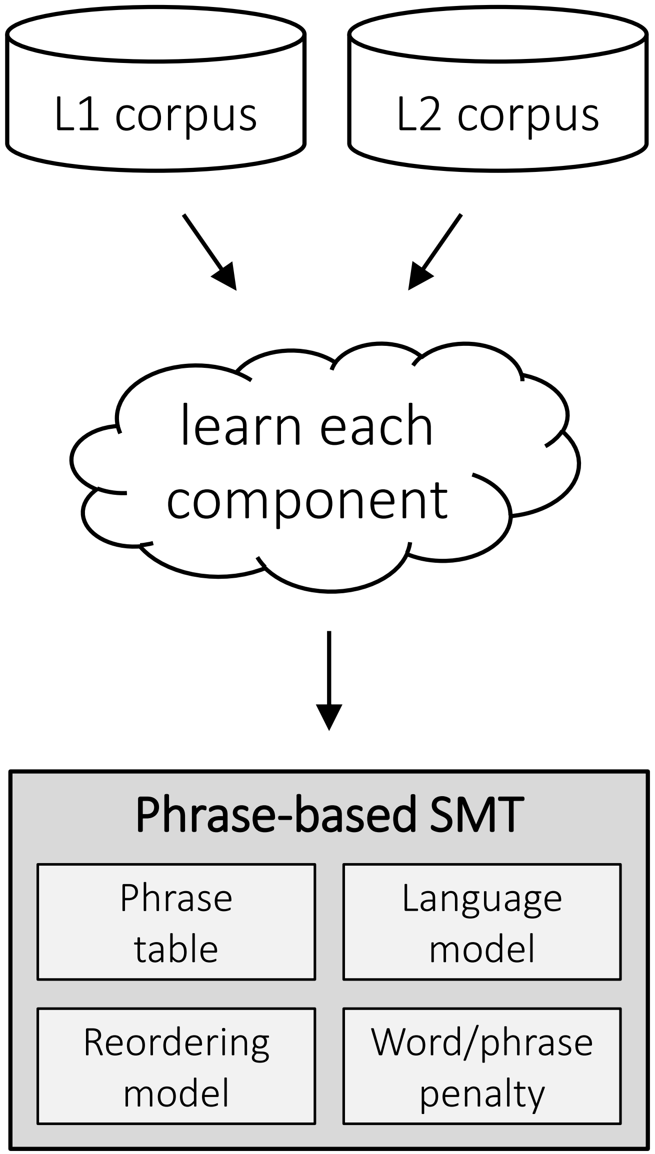
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - Lexical reordering model
- Word/phrase penalty
  - Fixed score to control the length of the output

Unsupervised phrase-based SMT



Learn components from monolingual corpora

- Phrase table
    - Direct/inverse translation probabilities
    - Direct/inverse lexical weightings
- Language model
    - N-gram frequency counts with back-off and smoothing
- Reordering model
    - Distortion model (distance based)
    - Lexical reordering model
- Word/phrase penalty EASY!!!
    - Fixed score to control the length of the output

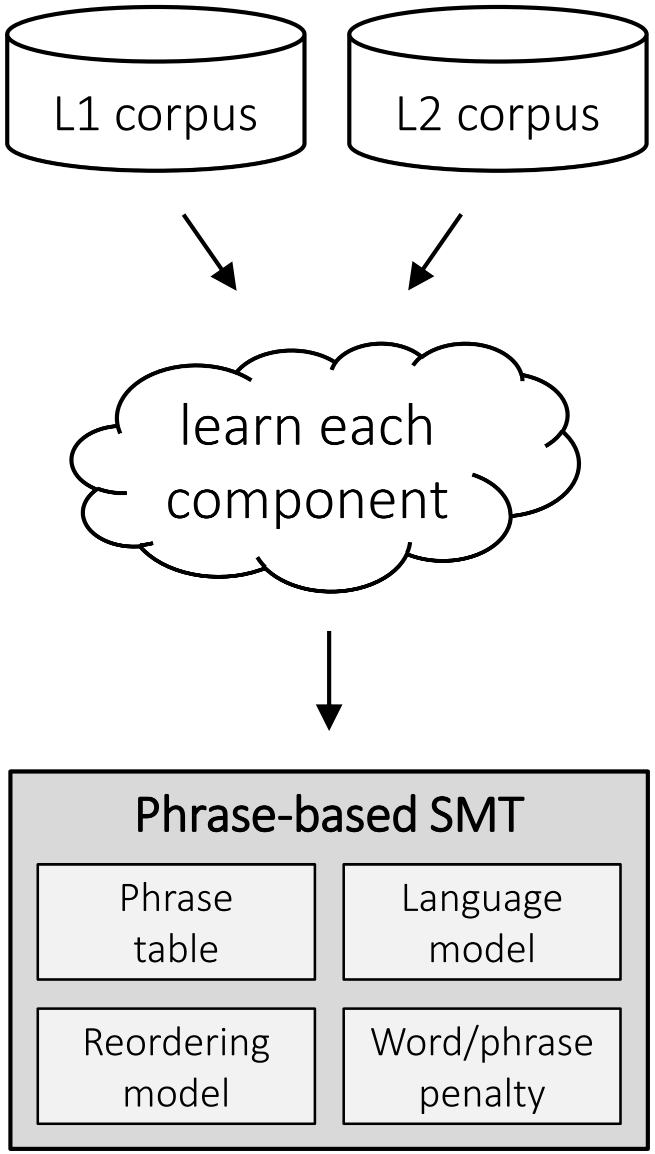


Unsupervised phrase-based SMT

Learn components from monolingual corpora

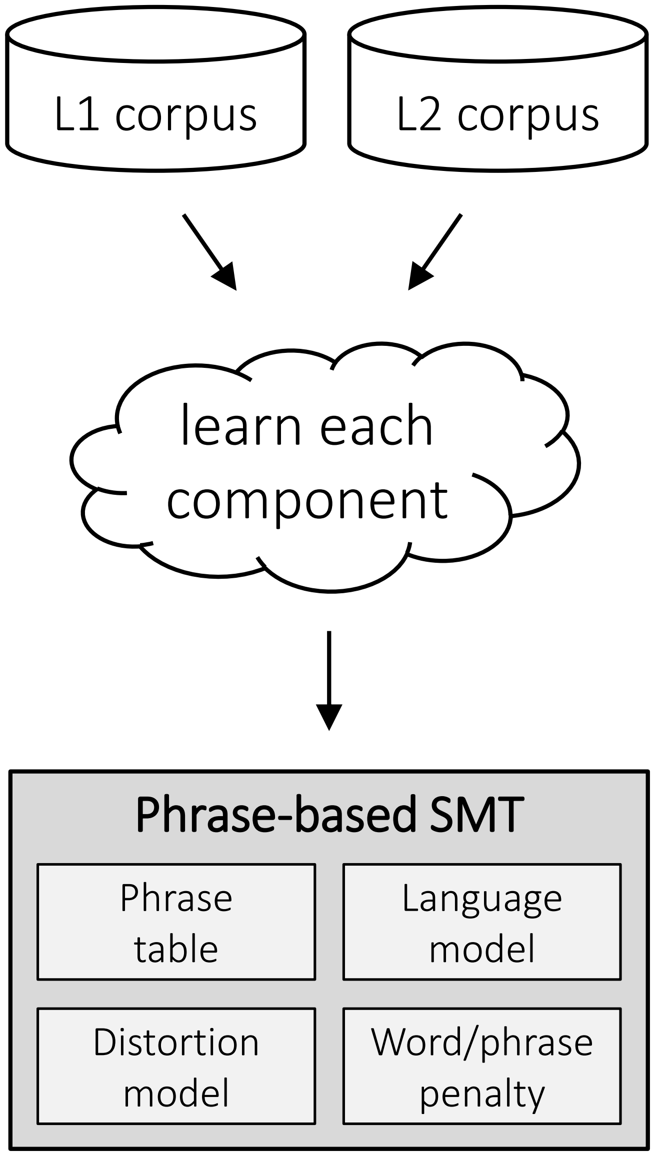
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - Lexical reordering model
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

Unsupervised phrase-based SMT



Learn components from monolingual corpora

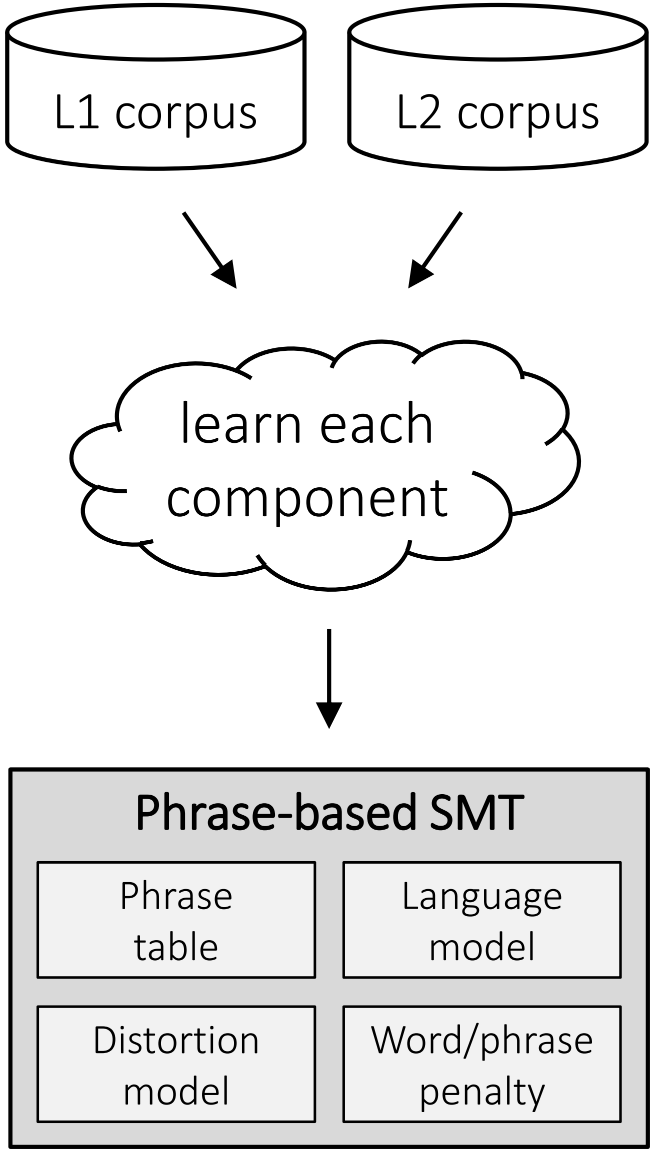
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

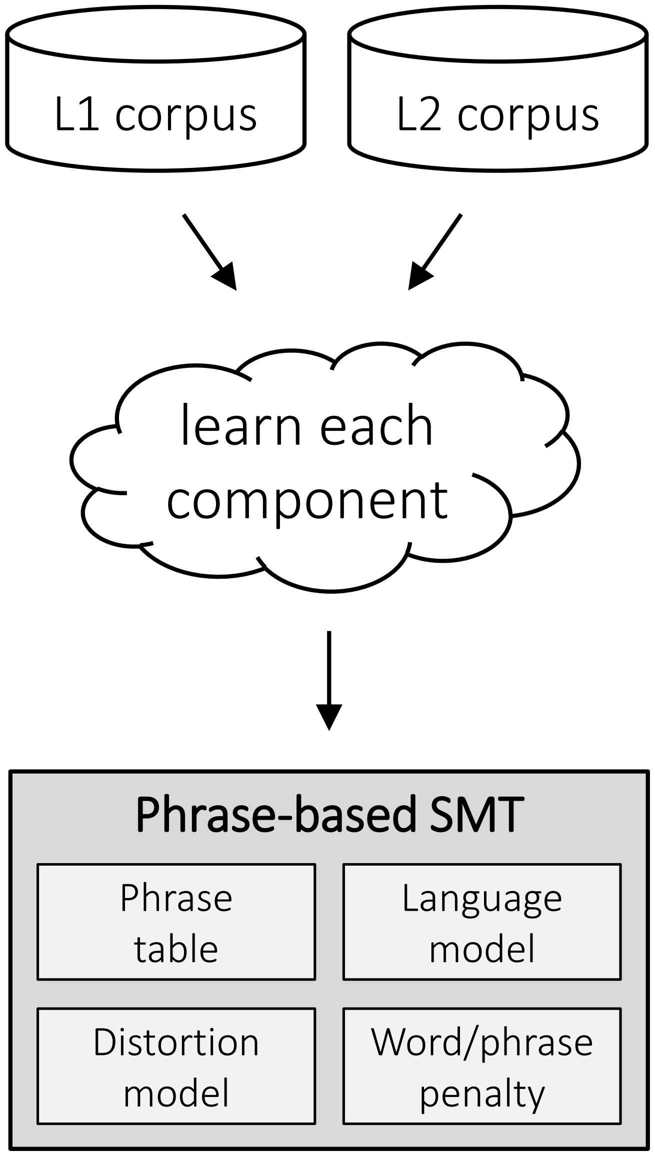
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

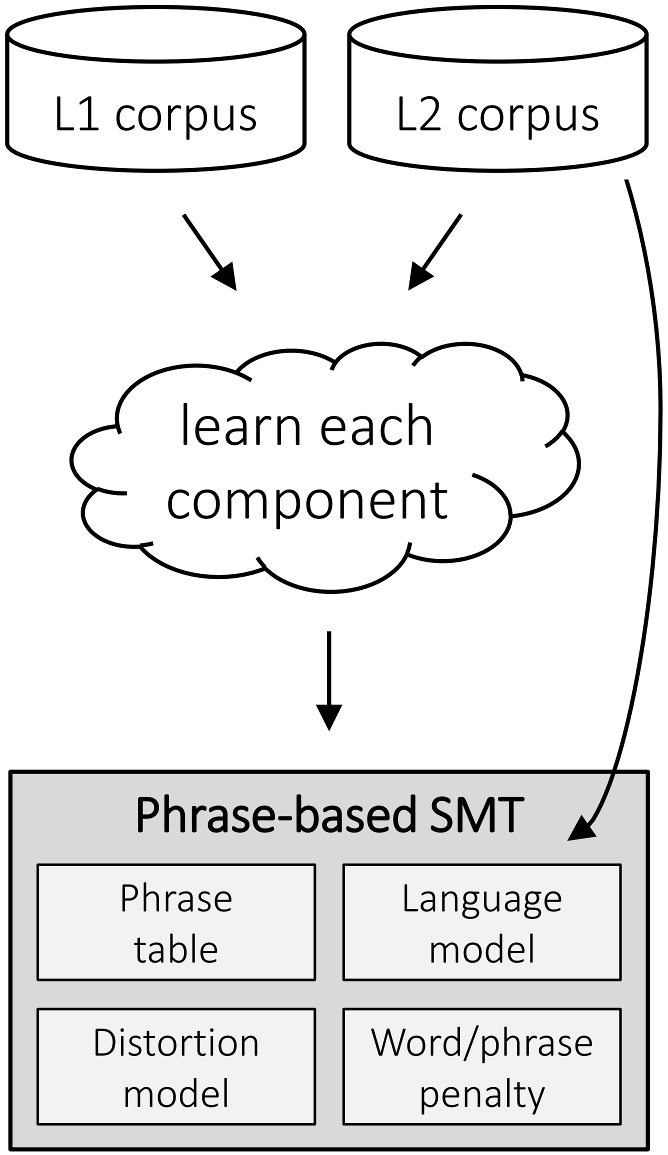
- Phrase table
    - Direct/inverse translation probabilities
    - Direct/inverse lexical weightings
- Language model
    - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
    - Distortion model (distance based)
    - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
    - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

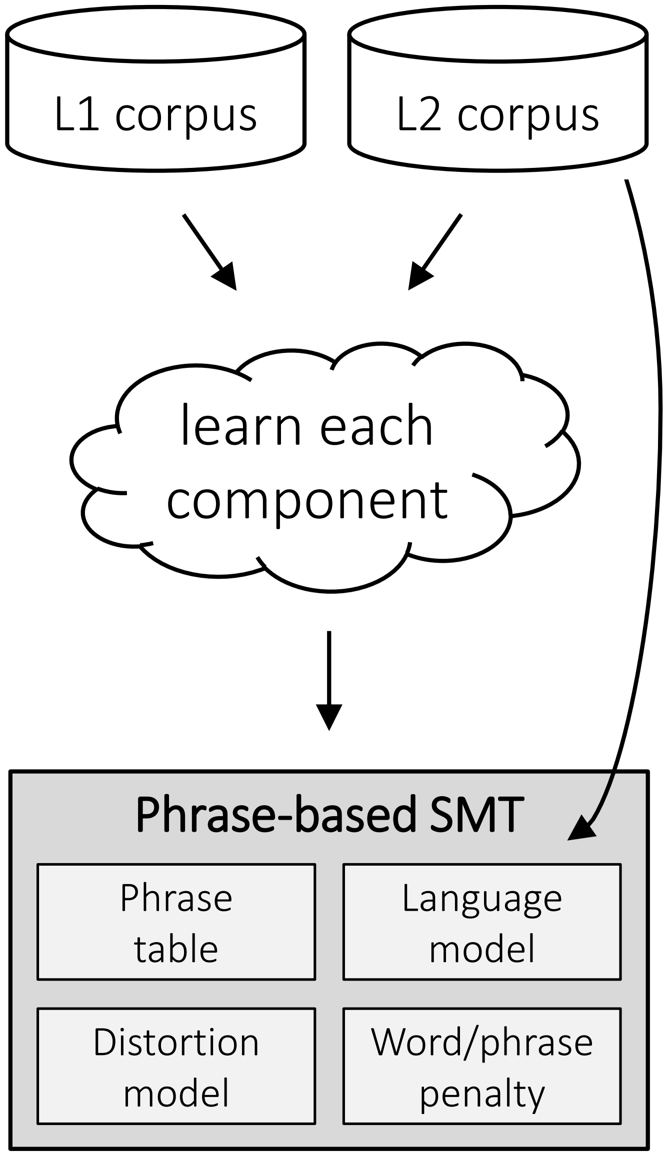
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

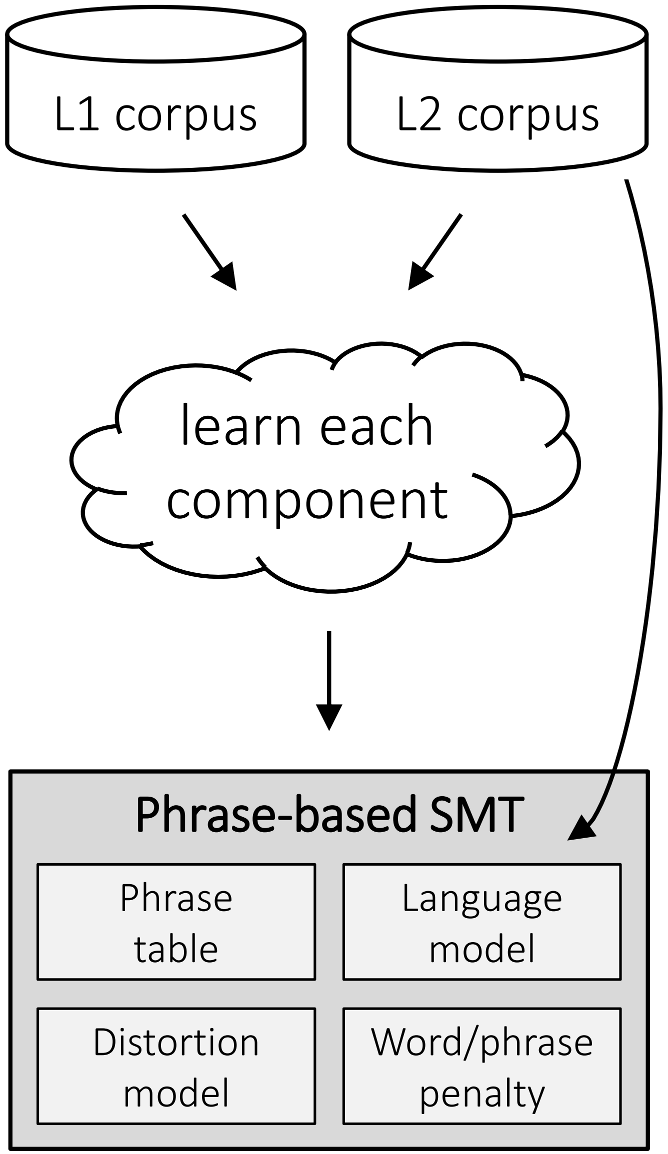


Unsupervised phrase-based SMT

Learn components from monolingual corpora

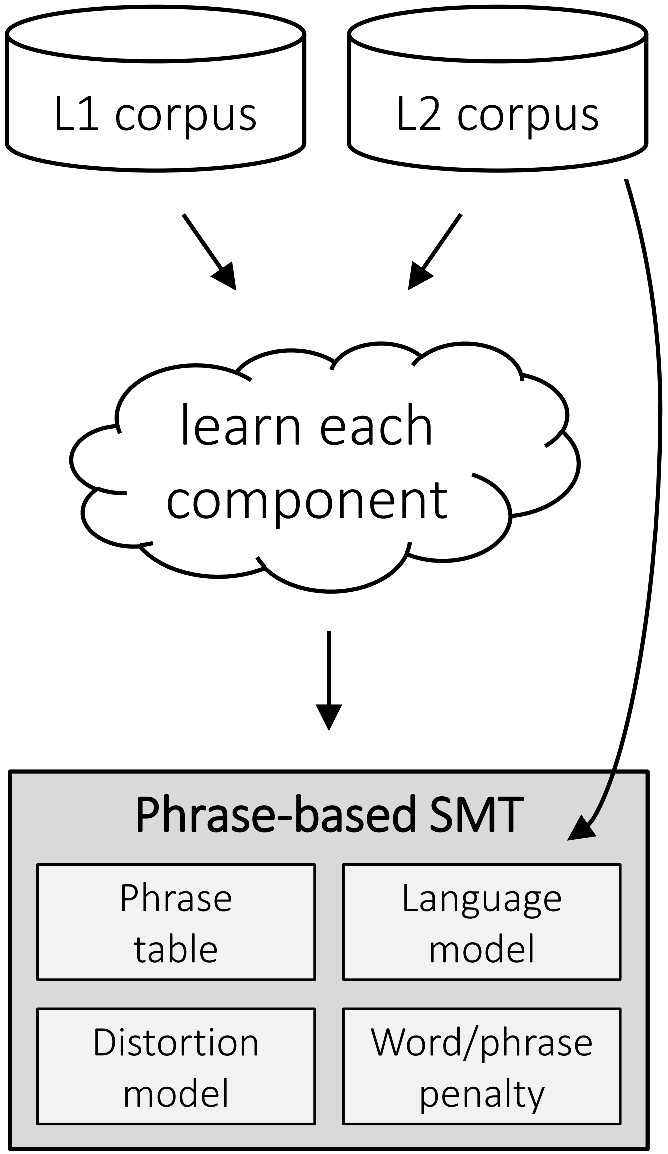
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

Unsupervised phrase-based SMT



Learn components from monolingual corpora

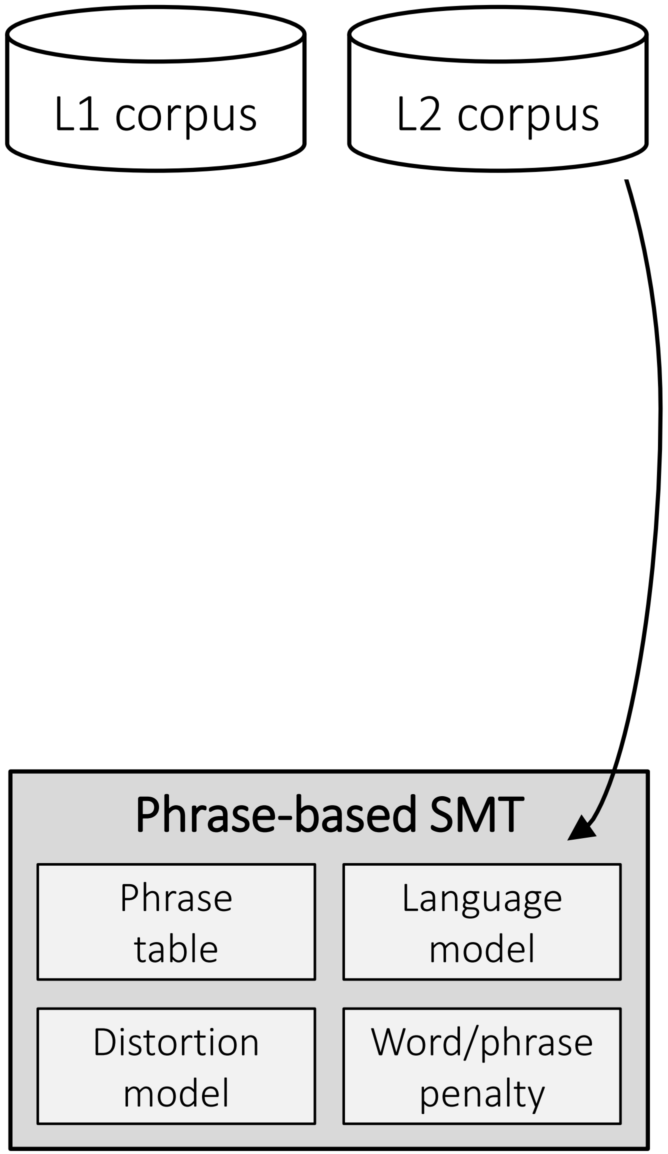
- Phrase table
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

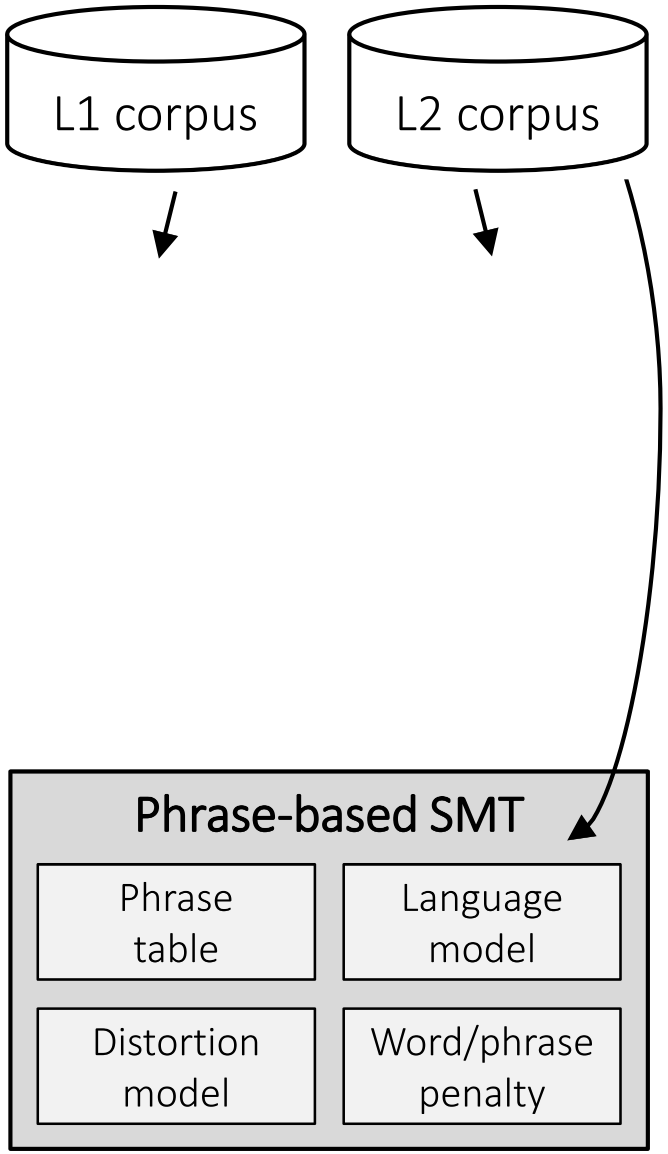
- Phrase table TRICKY…
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

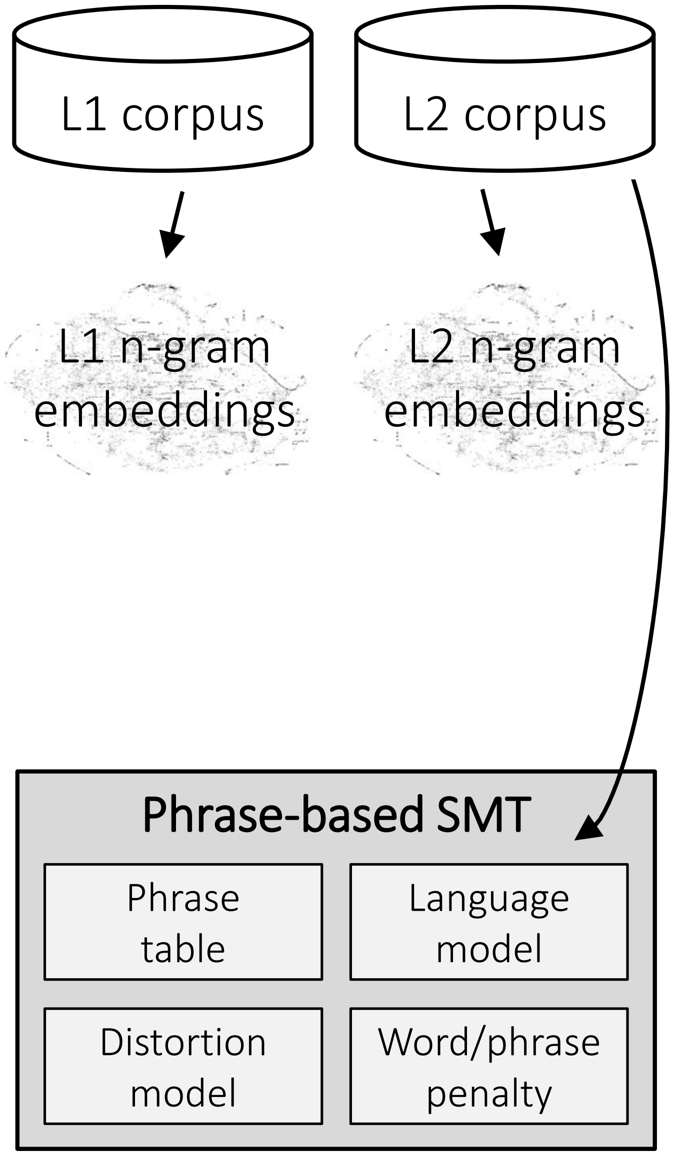
- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

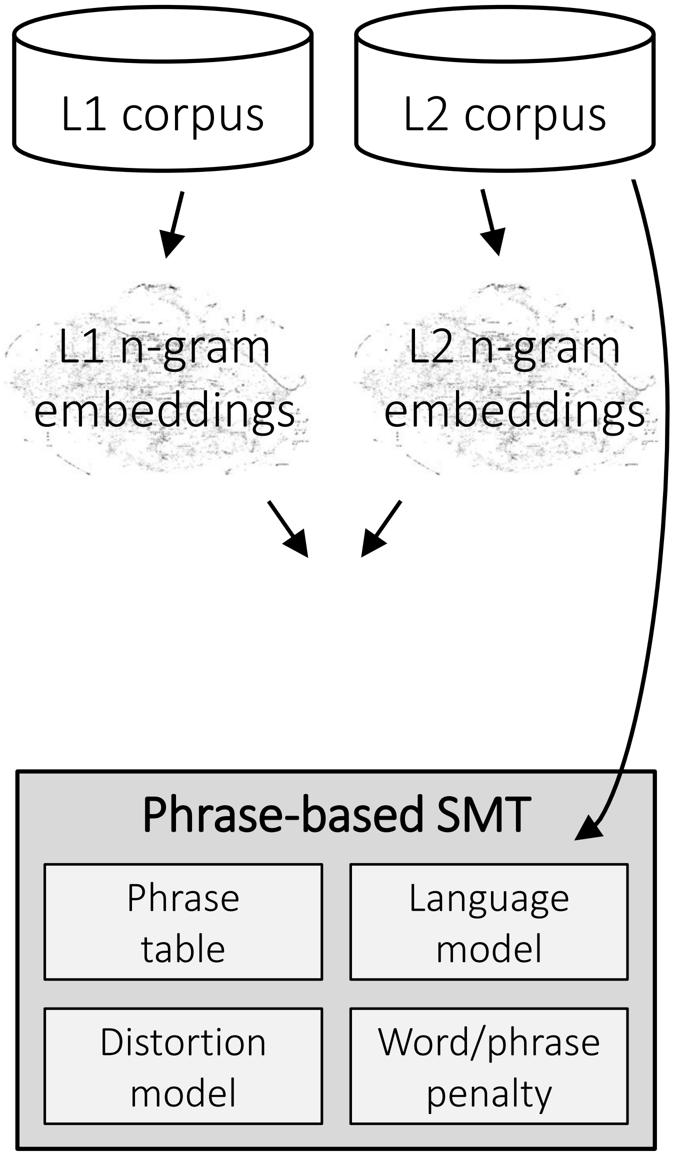
- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

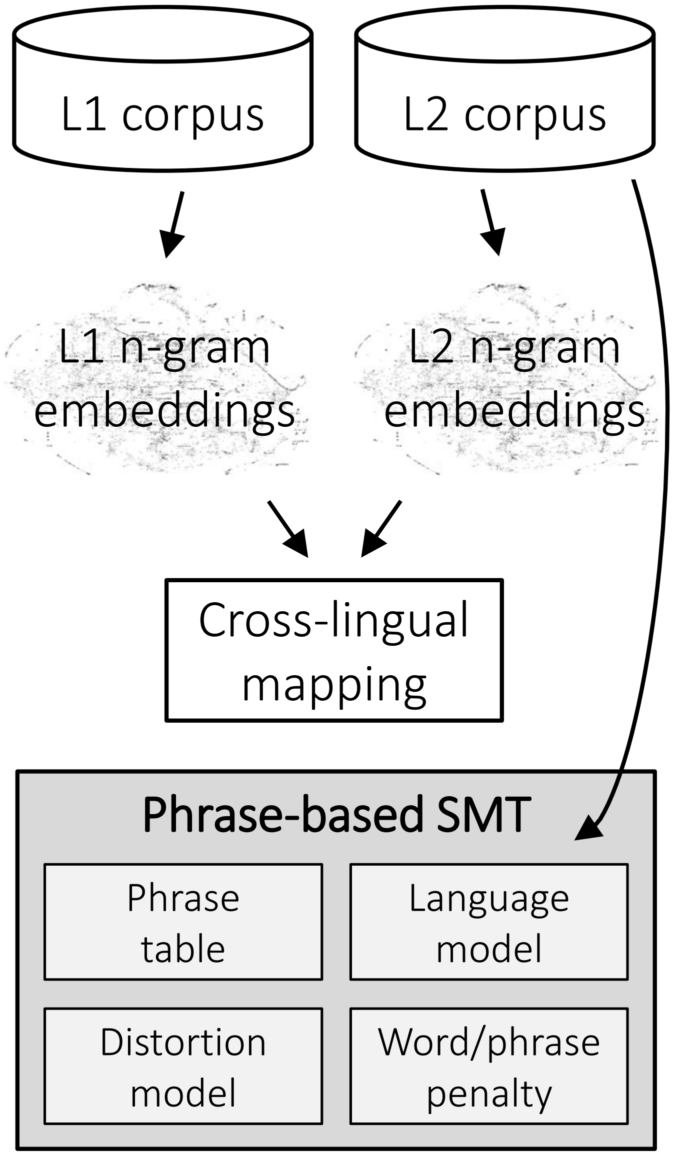
- Phrase table TRICKY...
    - Direct/inverse translation probabilities
    - Direct/inverse lexical weightings
- Language model EASY!!!
    - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
    - Distortion model (distance based)
    - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
    - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

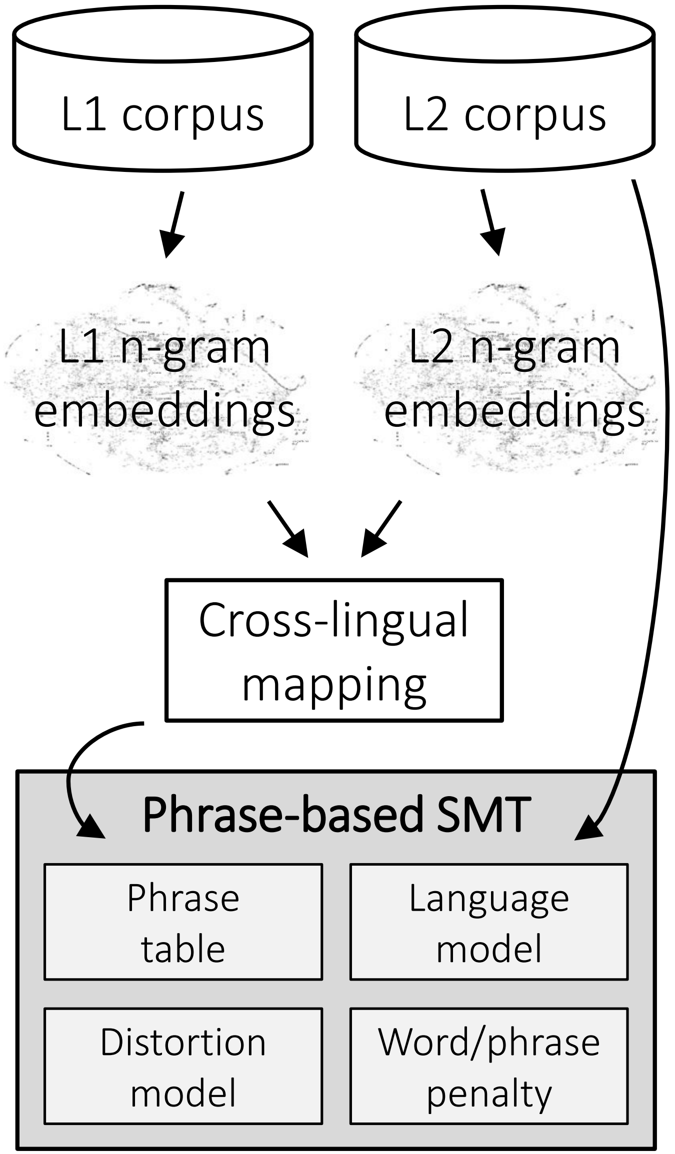
- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

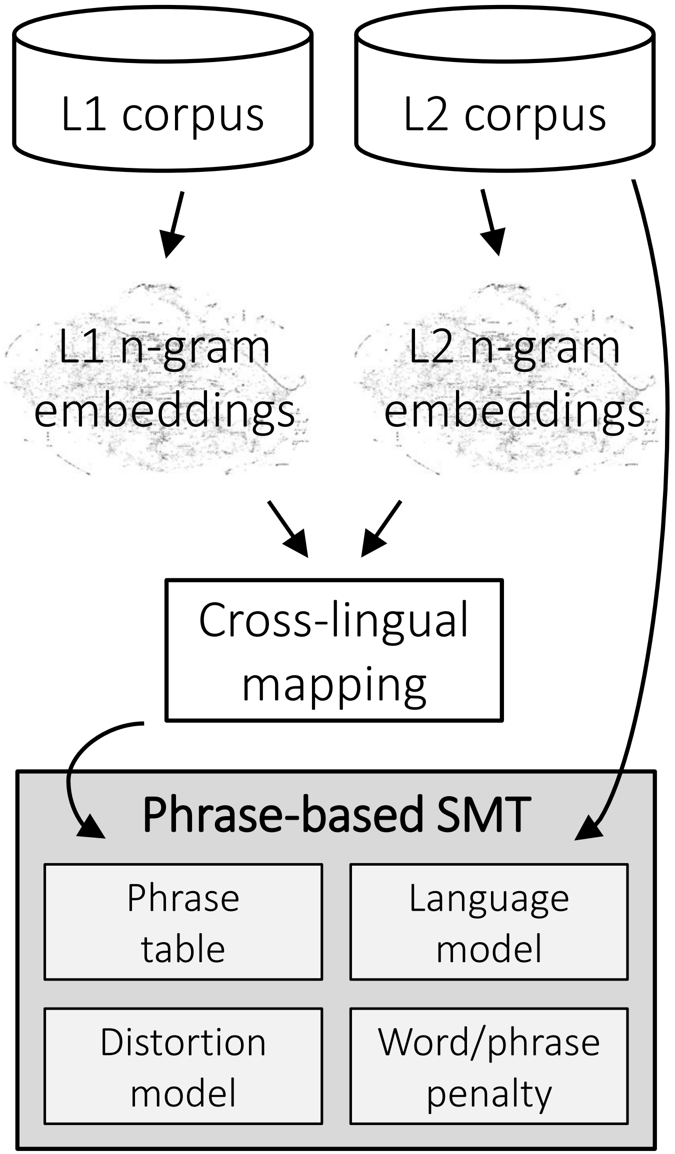
- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

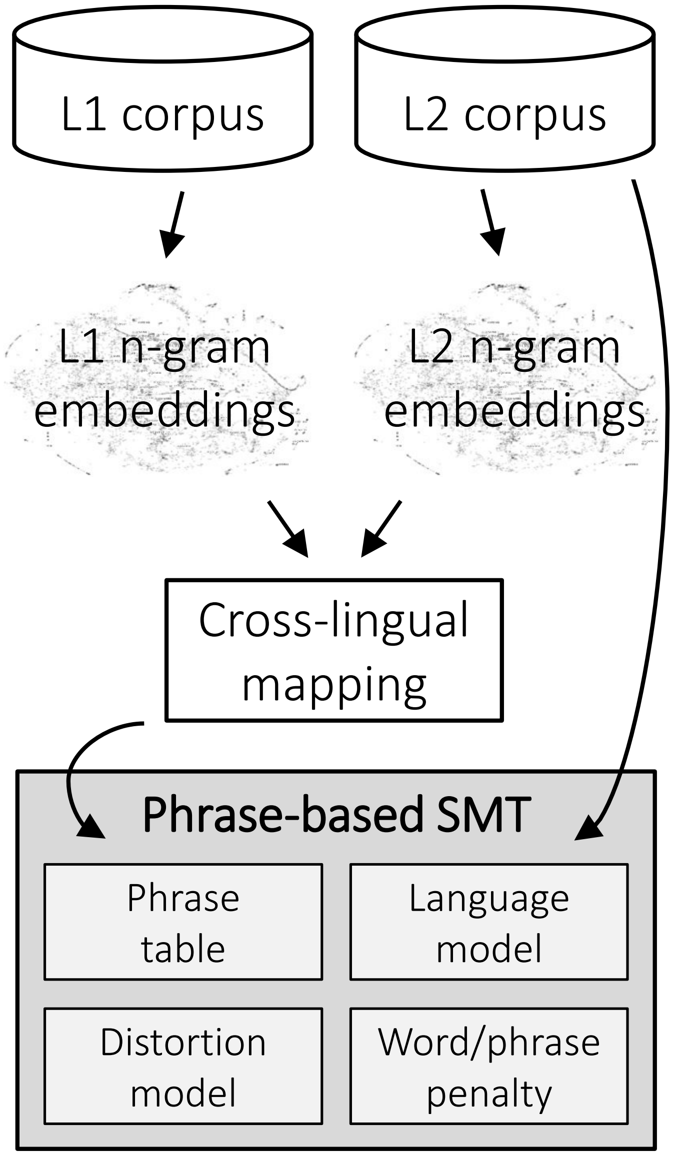
- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

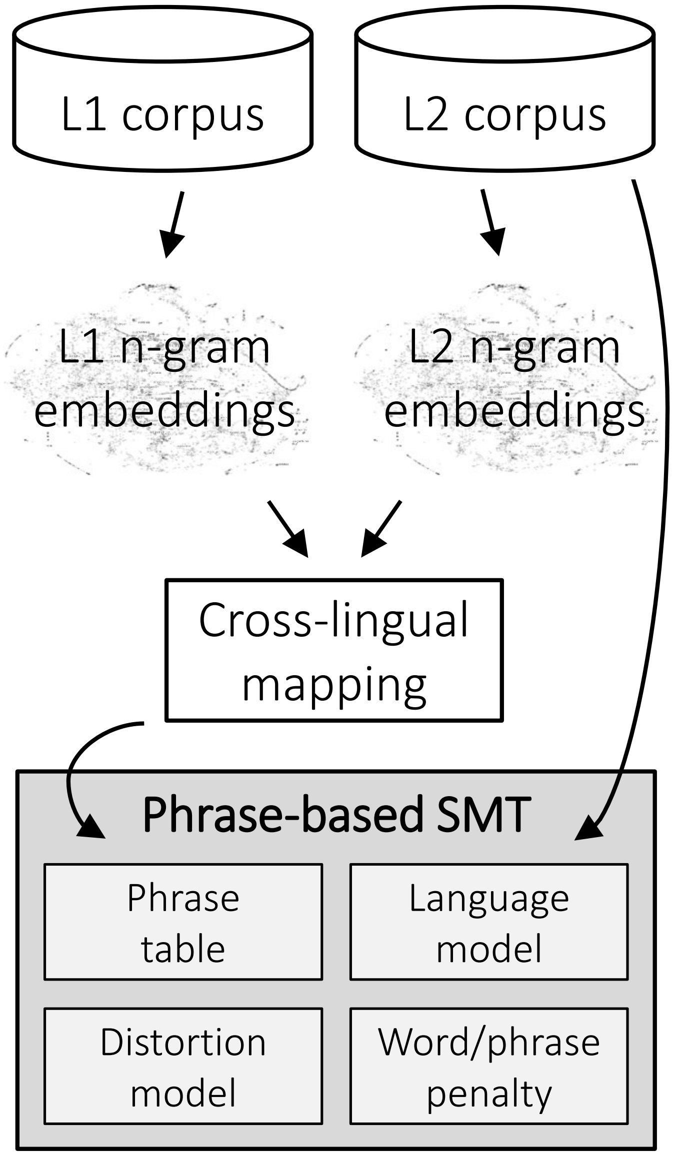


Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

I will go to New York by plane .

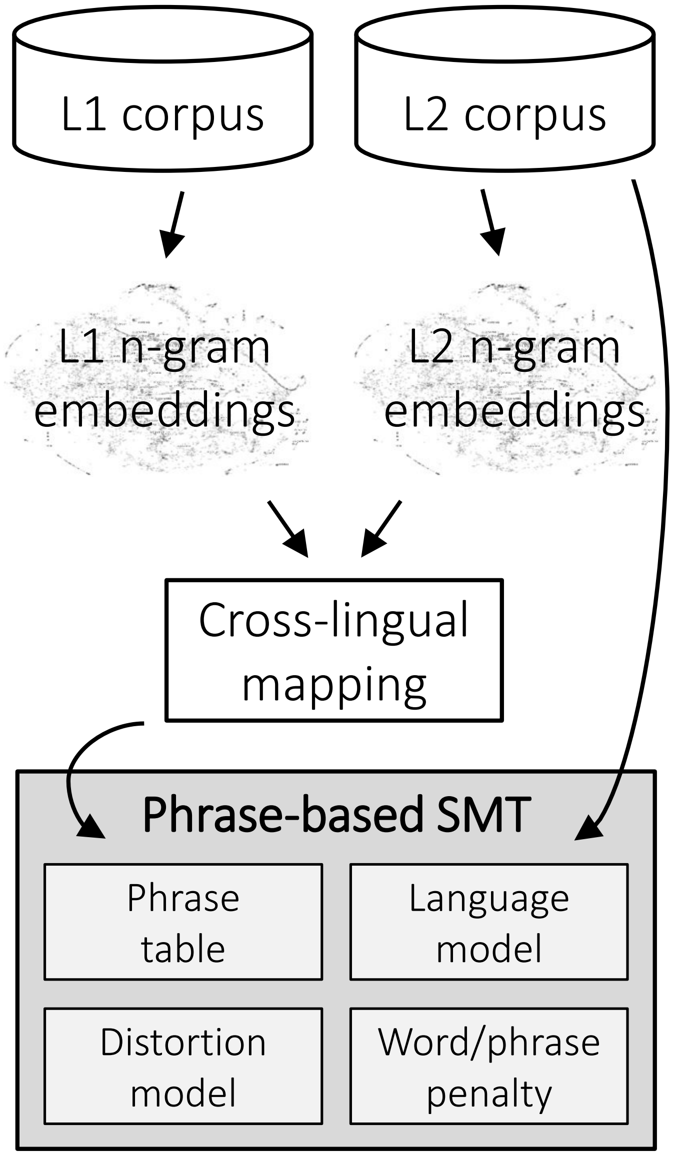


Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

I will $\underset{w}{\underline{\textit{go}}}$ *to New York by plane .*



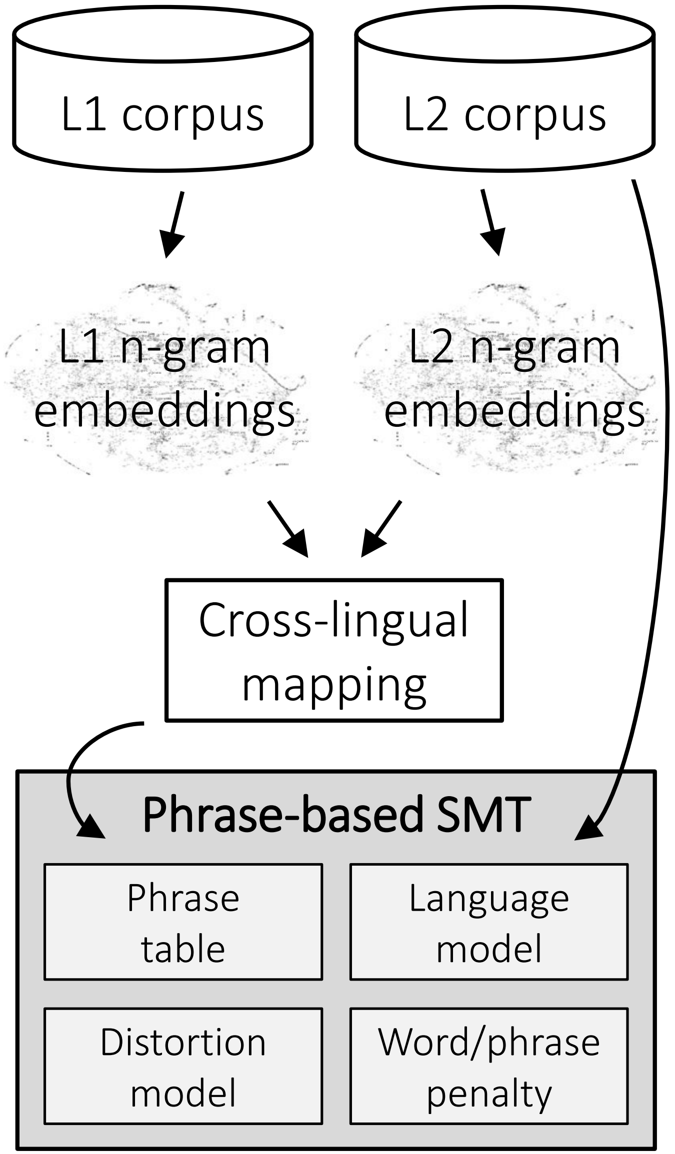
Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

I will go to New York by plane .

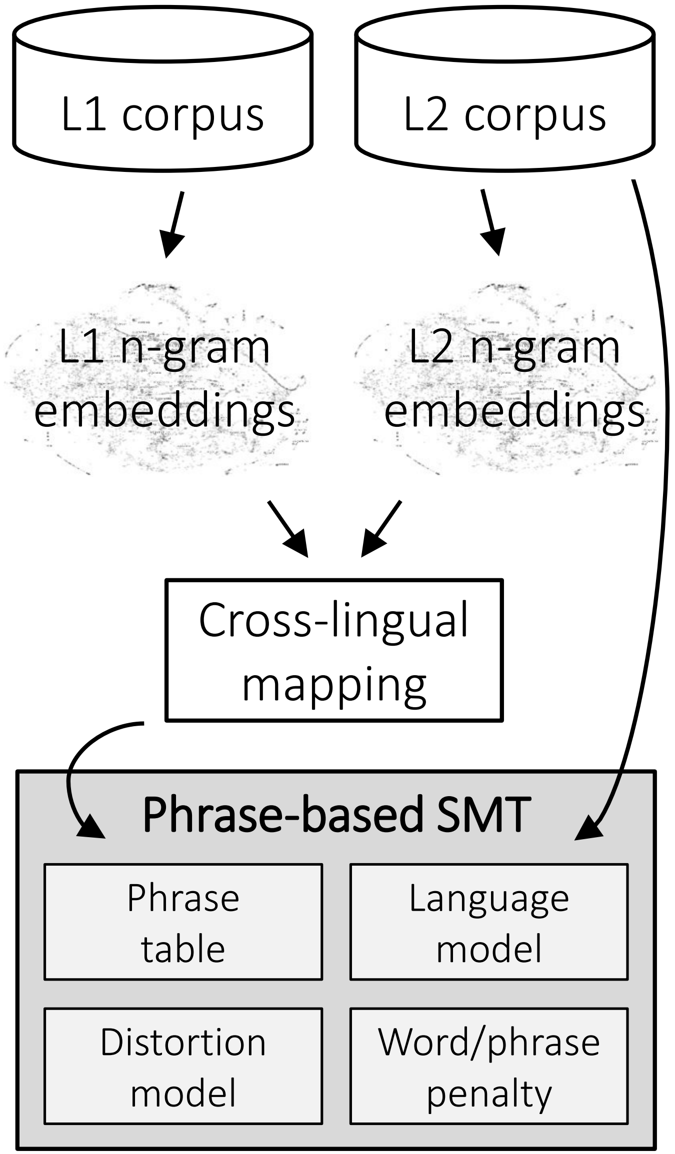
w



Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



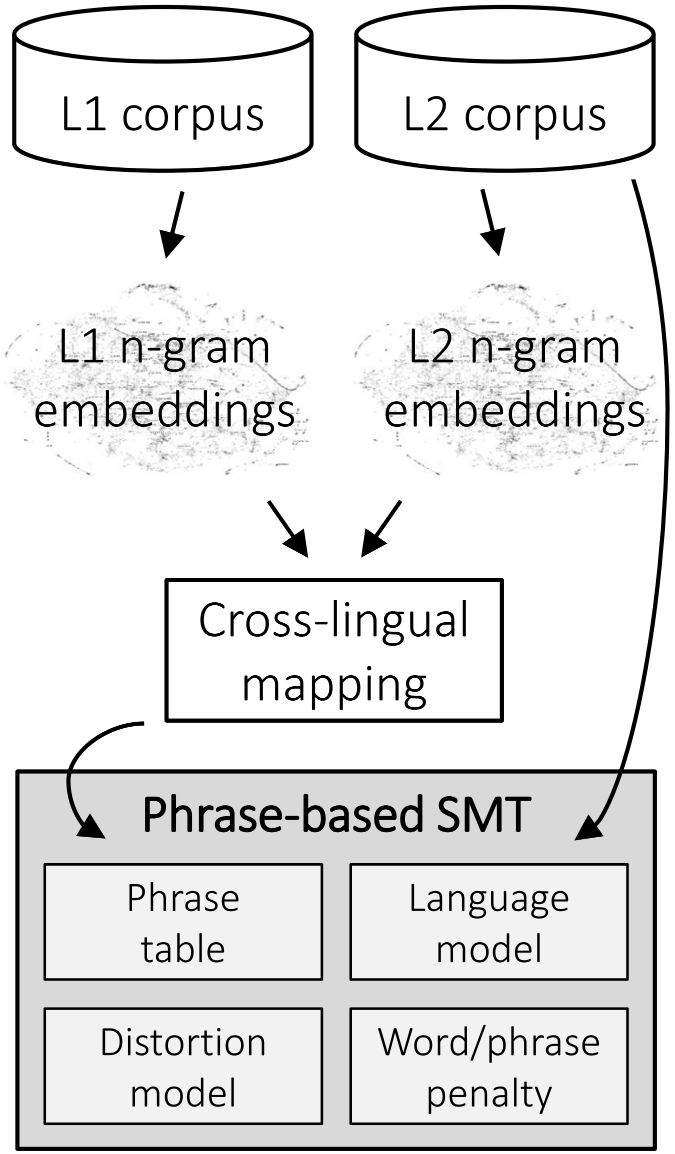
Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table **TRICKY…**
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model **EASY!!!**
  - N-gram frequency counts with back-off and smoothing
- Reordering model **EASY!!!**
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty **EASY!!!**
  - Fixed score to control the length of the output

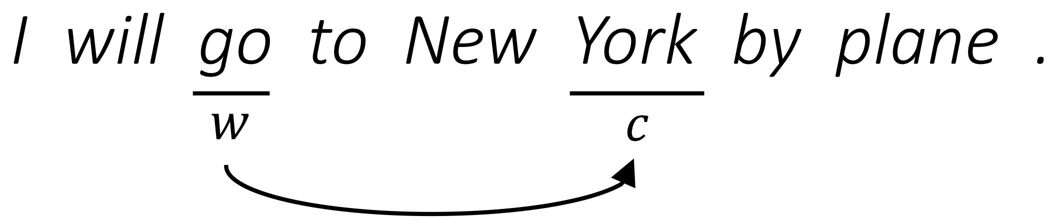
I will $\underset{w}{\underline{\textit{go}}}$ *to* $\underset{c}{\underline{\textit{New}}}$ *York by plane .*

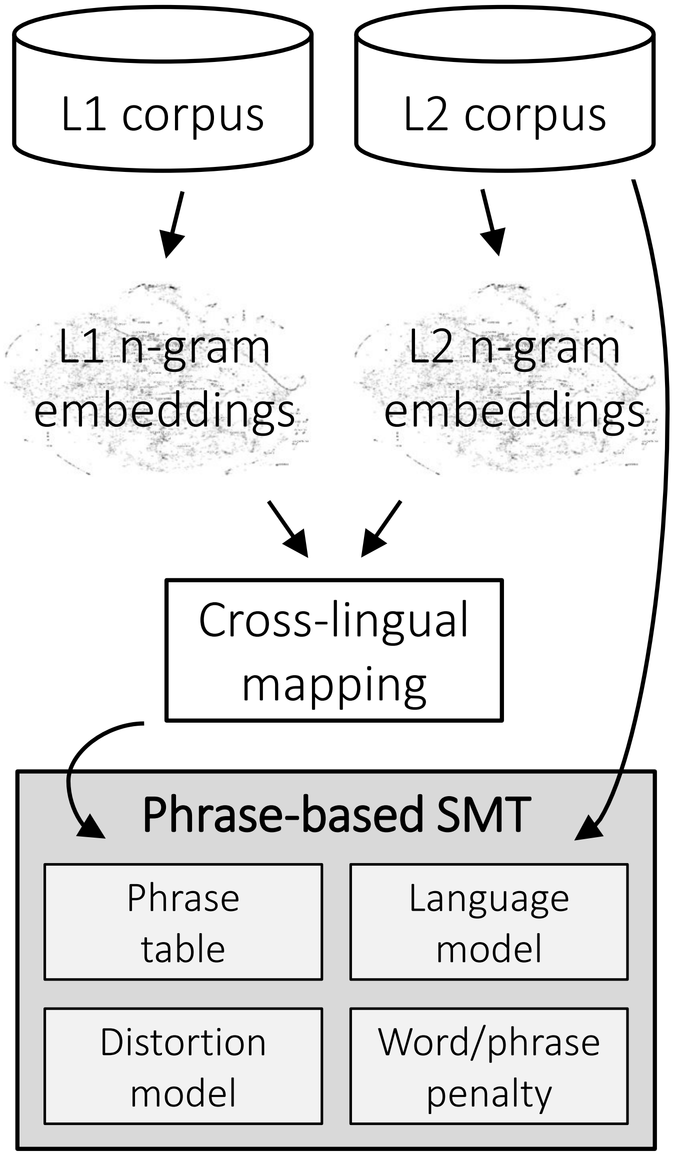
Unsupervised phrase-based SMT



Learn components from monolingual corpora

- Phrase table TRICKY…
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output





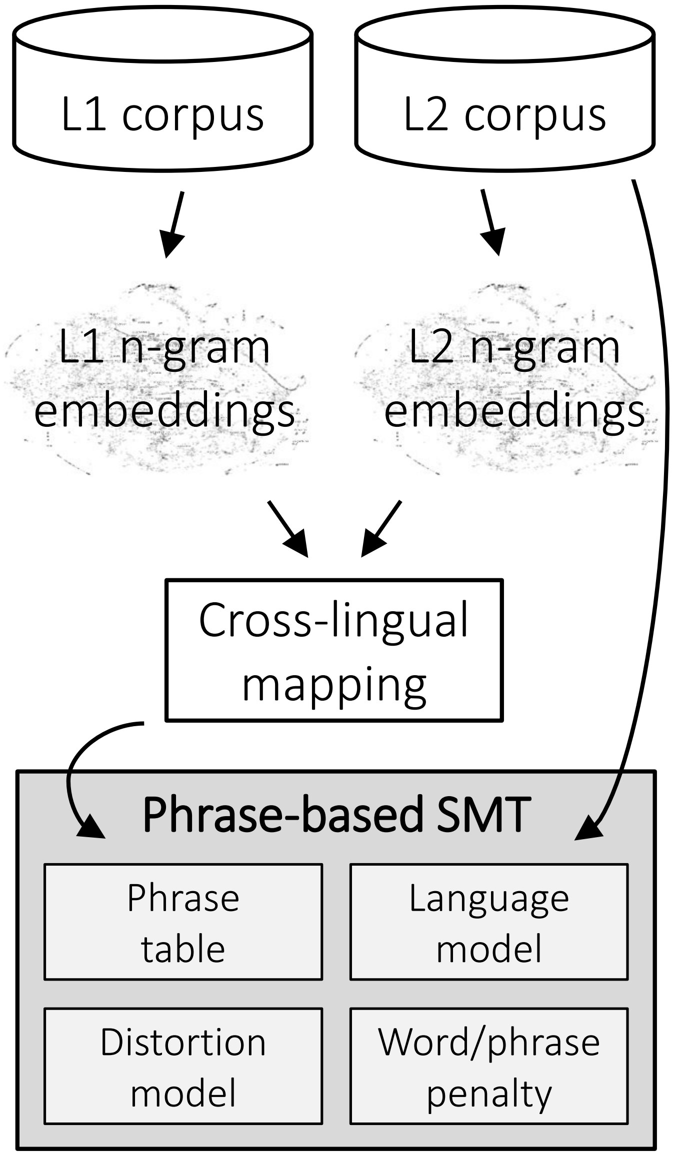
Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

I will go to New York by plane .

p c

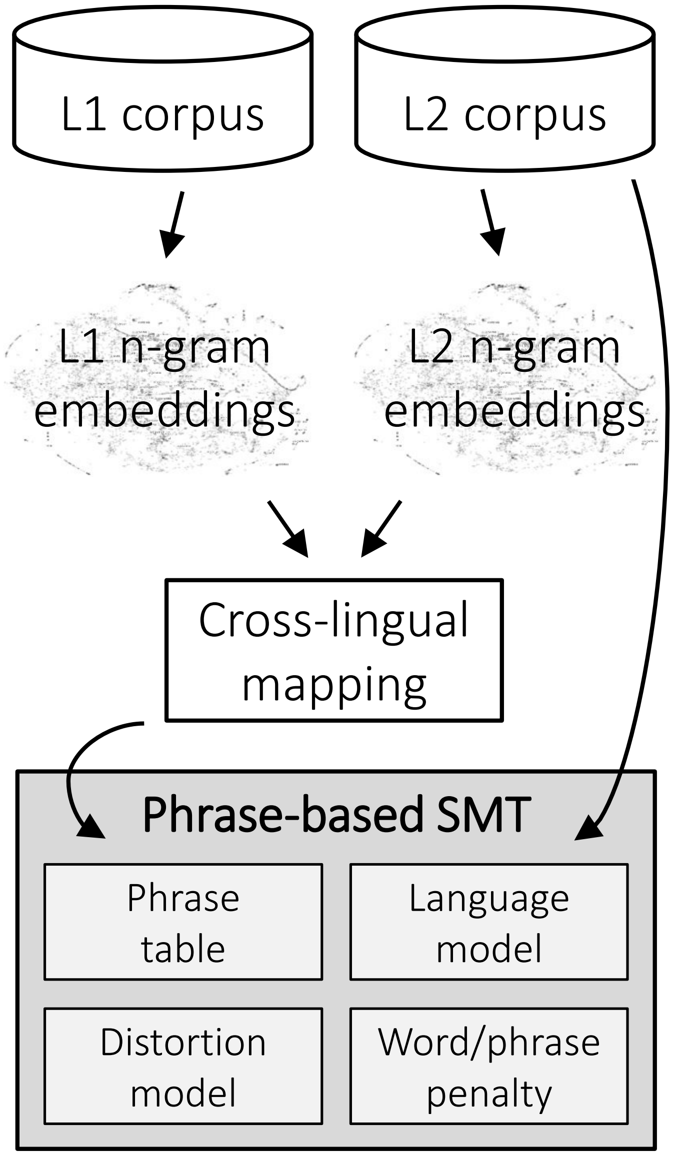


Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table TRICKY…
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

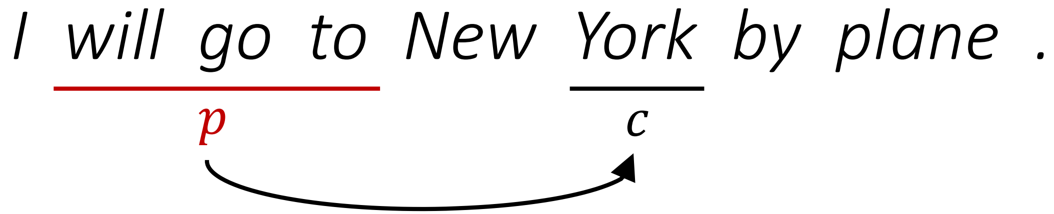
I $\underbrace{\textit{will go}}_{p}$ *to New* $\underbrace{\textit{York}}_{c}$ *by plane .*

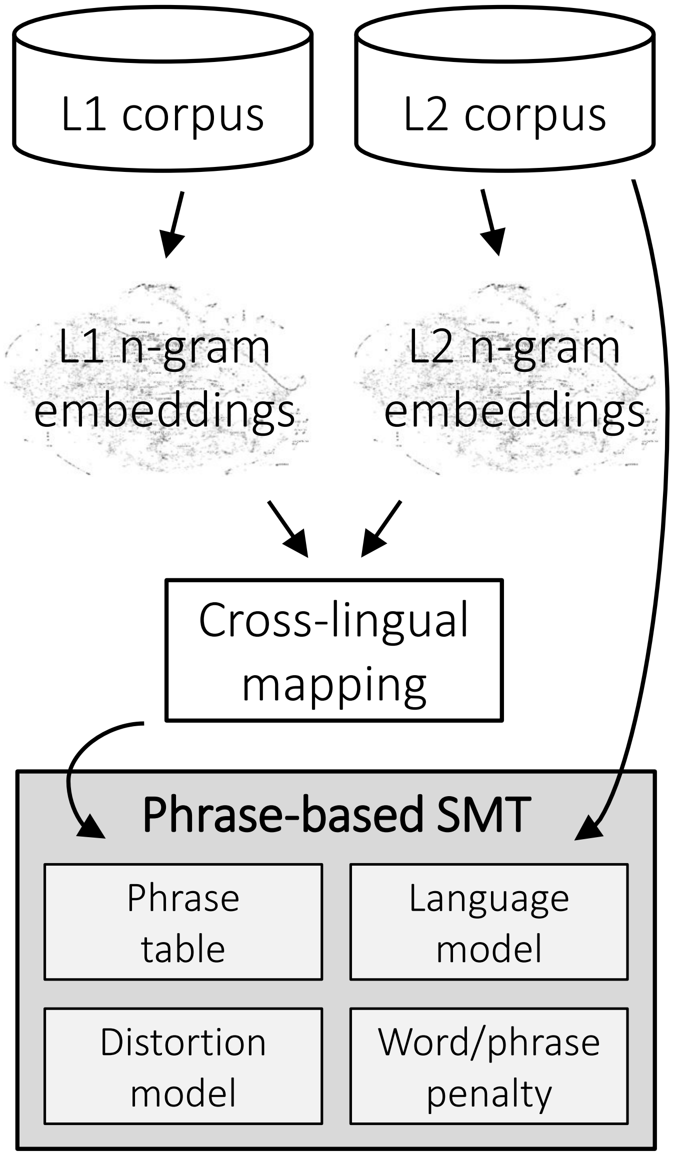


Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

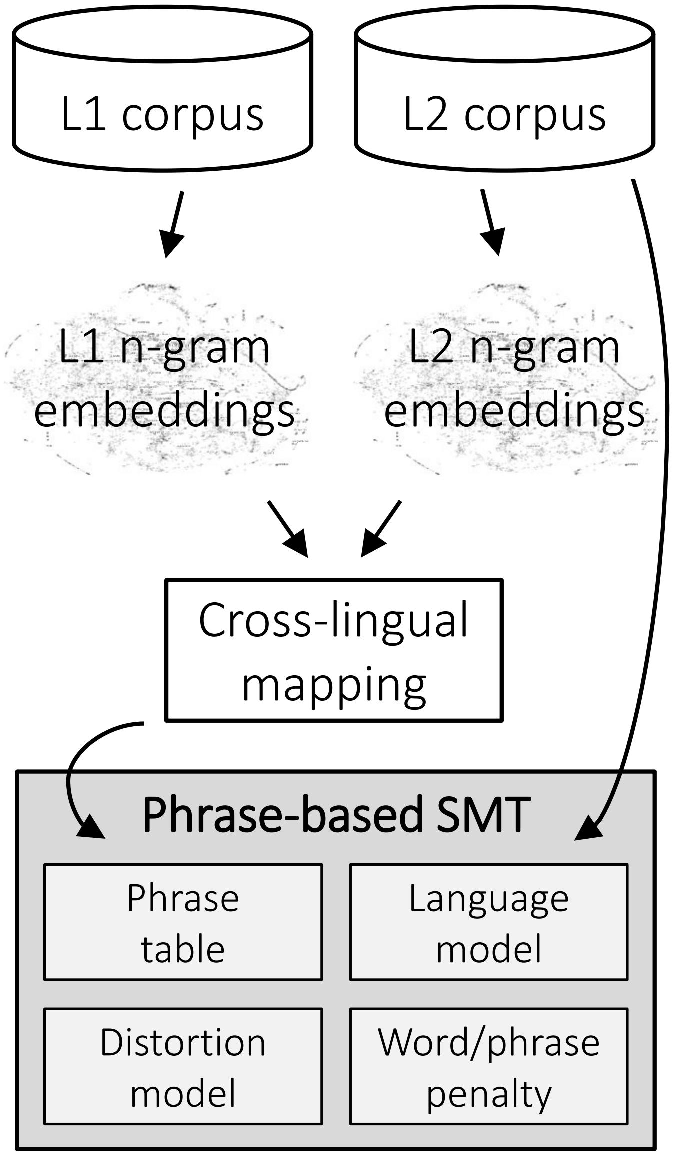




Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table TRICKY...
    - Direct/inverse translation probabilities
    - Direct/inverse lexical weightings
- Language model EASY!!!
    - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
    - Distortion model (distance based)
    - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
    - Fixed score to control the length of the output

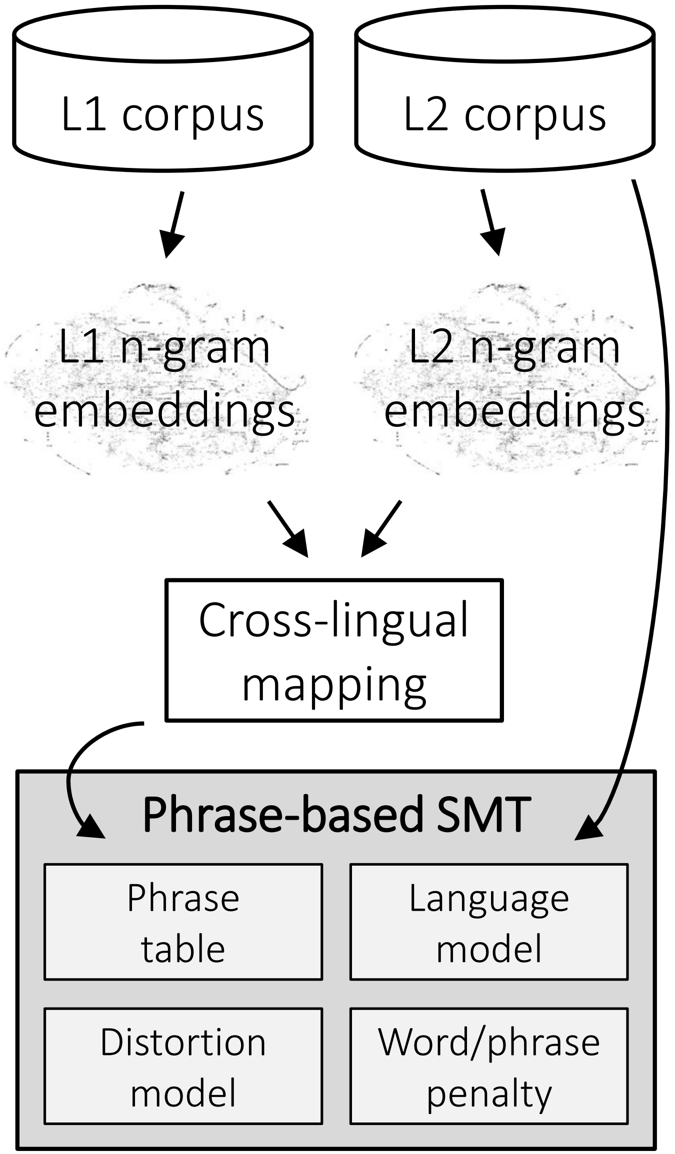


Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

For each $\bar{e}$, estimate $\phi(\bar{f}|\bar{e})$ for 100-NN:



Unsupervised phrase-based SMT

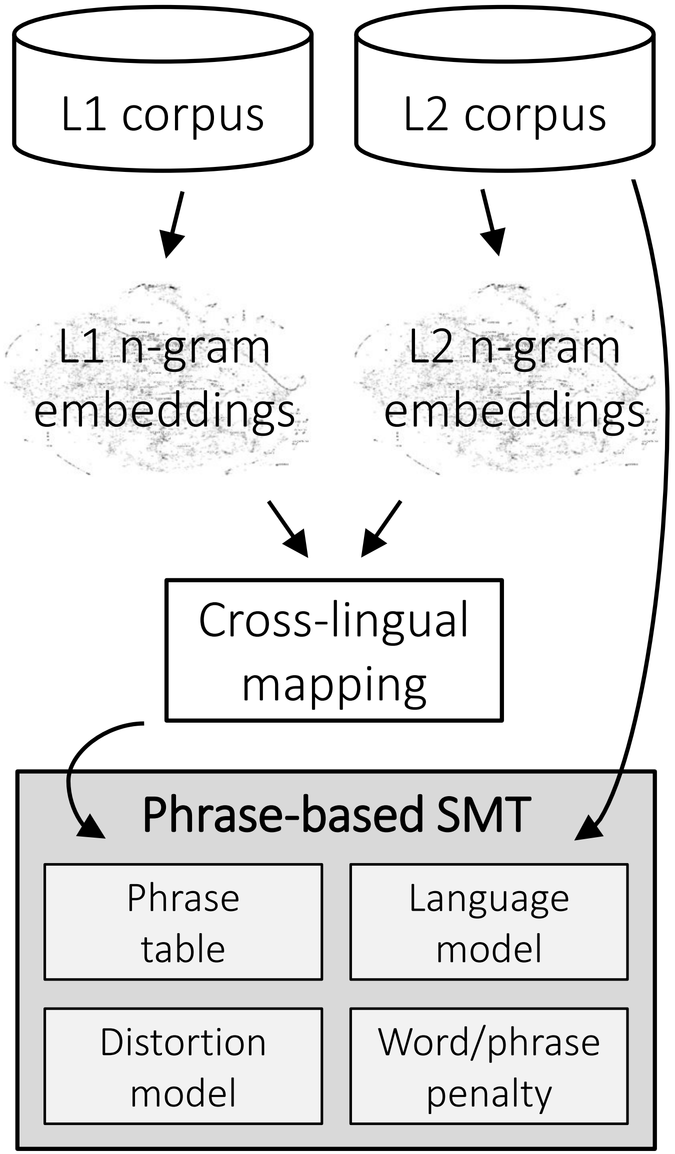
Learn components from monolingual corpora

- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

For each $\bar{e}$, estimate $\phi(\bar{f}|\bar{e})$ for 100-NN:

$$\phi(\bar{f}|\bar{e}) = \frac{e^{\cos(\bar{e},\bar{f})/\tau}}{\sum_{\overline{f'}} e^{\cos(\bar{e},\overline{f'})/\tau}}$$

Unsupervised phrase-based SMT



Learn components from monolingual corpora

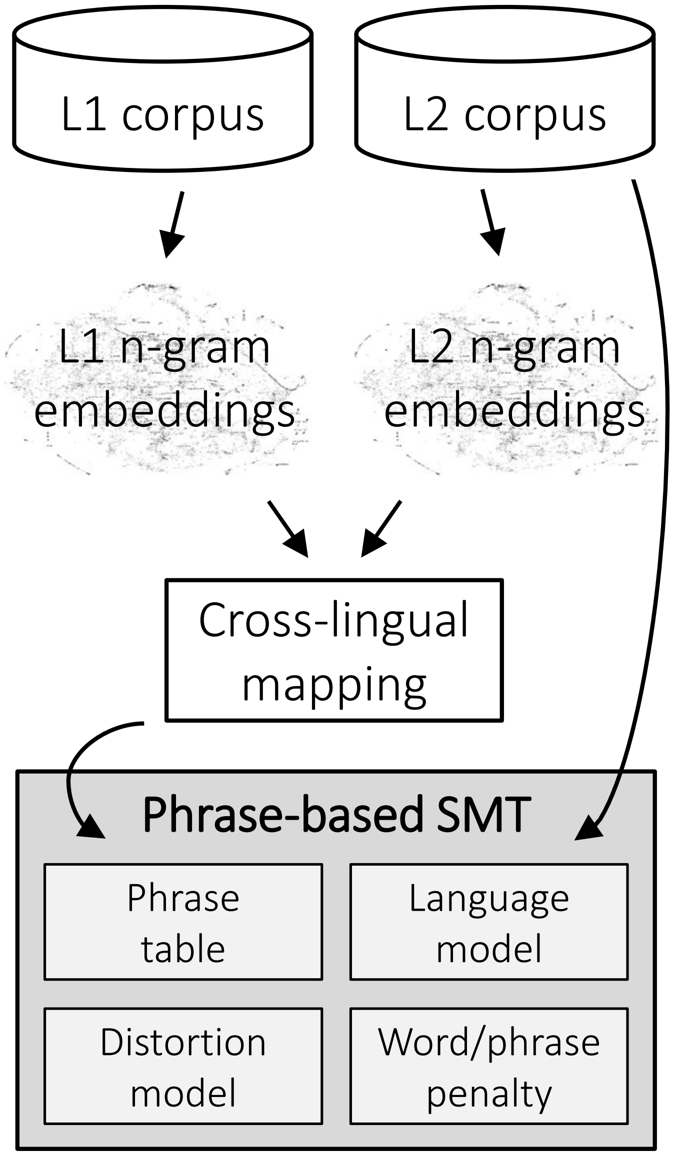
- Phrase table TRICKY…
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

For each $\bar{e}$, estimate $\phi(\bar{f}|\bar{e})$ for 100-NN:

$$\phi(\bar{f}|\bar{e}) = \frac{e^{\cos(\bar{e},\bar{f})/\tau}}{\sum_{\bar{f}'} e^{\cos(\bar{e},\bar{f}')/\tau}}$$

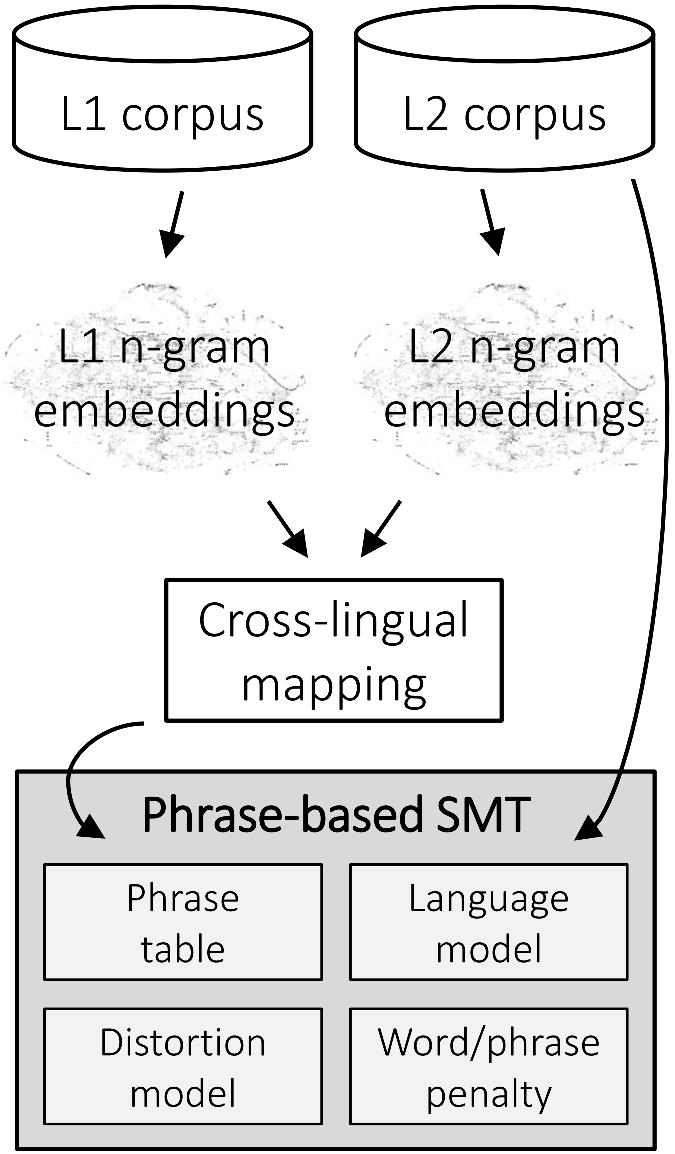
$$\min_{\tau} \sum_{\bar{f}} \log \phi\left(\bar{f}|\mathrm{NN}_{\bar{e}}(\bar{f})\right) + \sum_{\bar{e}} \log \phi\left(\bar{e}|\mathrm{NN}_{\bar{f}}(\bar{e})\right)$$

Unsupervised phrase-based SMT



Learn components from monolingual corpora

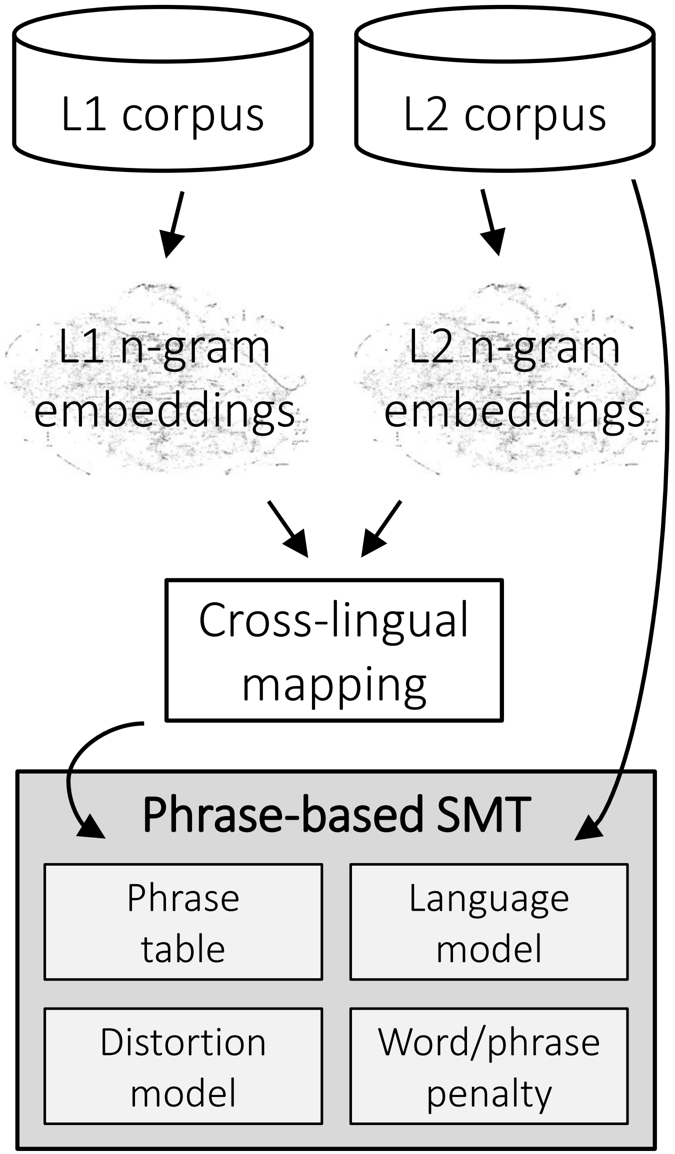
- Phrase table TRICKY...
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output



Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table ~~TRICKY…~~
    - Direct/inverse translation probabilities
    - Direct/inverse lexical weightings
- Language model EASY!!!
    - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
    - Distortion model (distance based)
    - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
    - Fixed score to control the length of the output

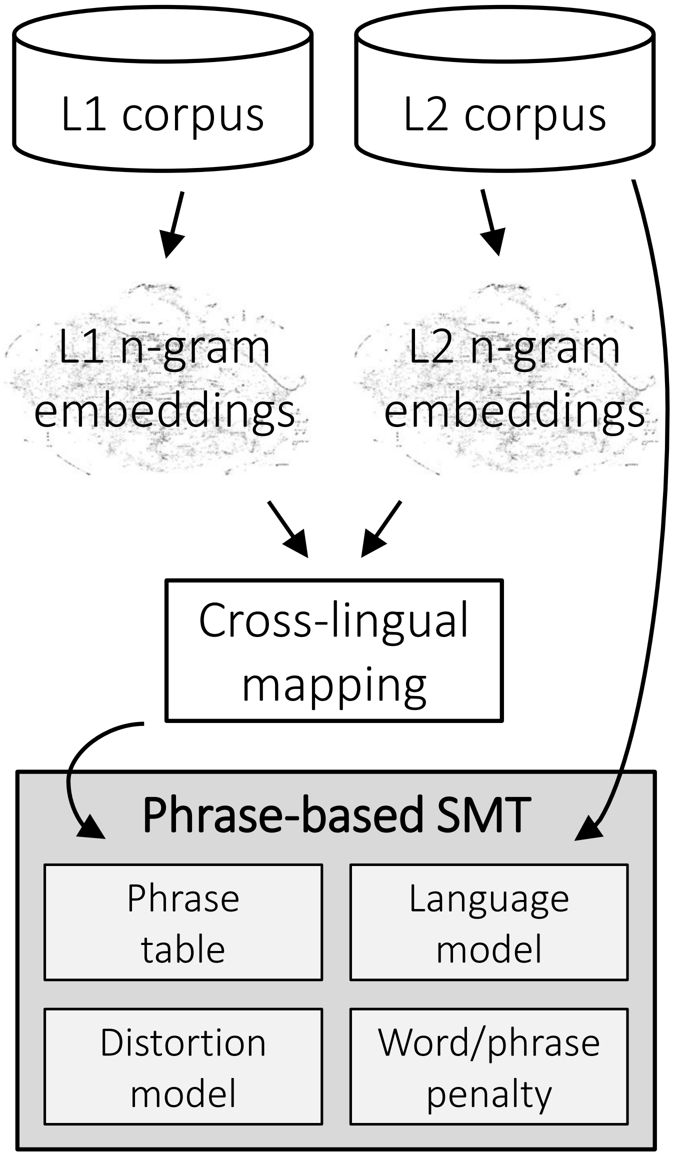


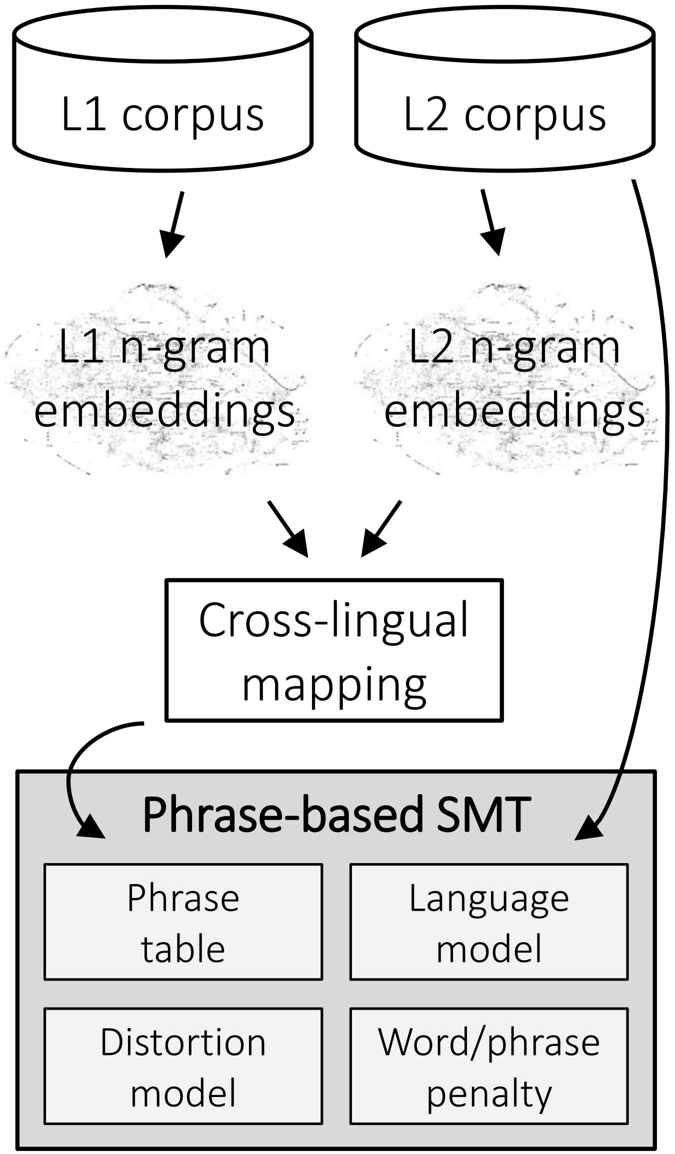
Unsupervised phrase-based SMT

Learn components from monolingual corpora

- Phrase table ~~TRICKY…~~ EASY!!!
  - Direct/inverse translation probabilities
  - Direct/inverse lexical weightings
- Language model EASY!!!
  - N-gram frequency counts with back-off and smoothing
- Reordering model EASY!!!
  - Distortion model (distance based)
  - ~~Lexical reordering model~~
- Word/phrase penalty EASY!!!
  - Fixed score to control the length of the output

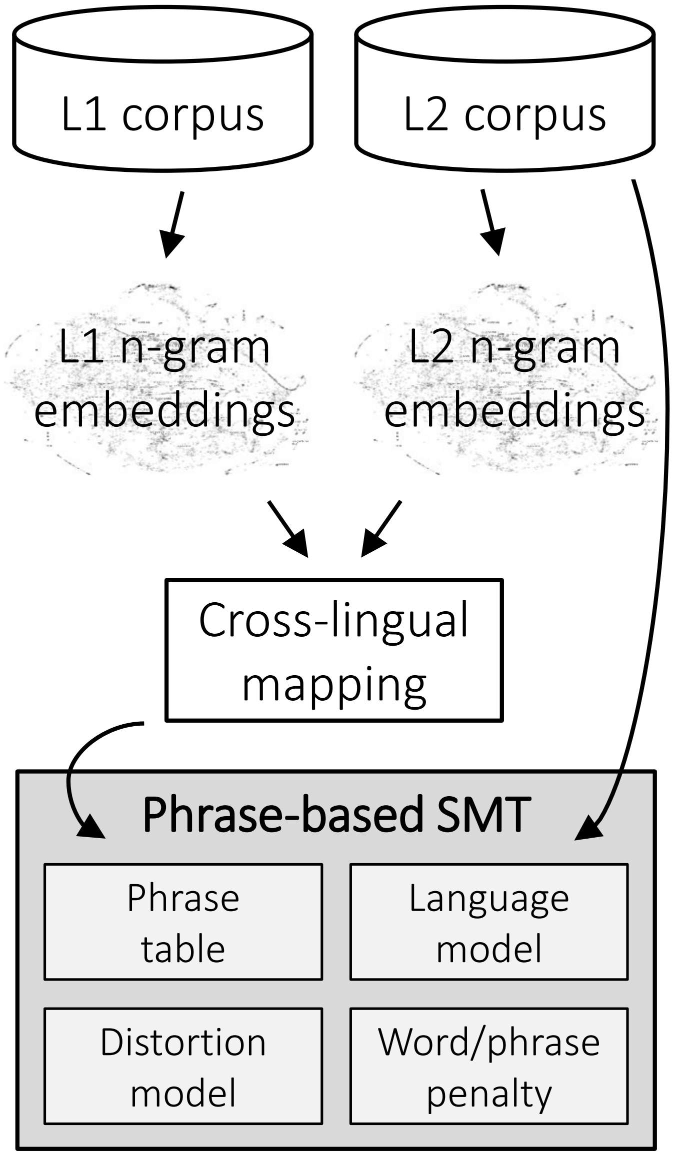
Unsupervised phrase-based SMT





Unsupervised phrase-based SMT

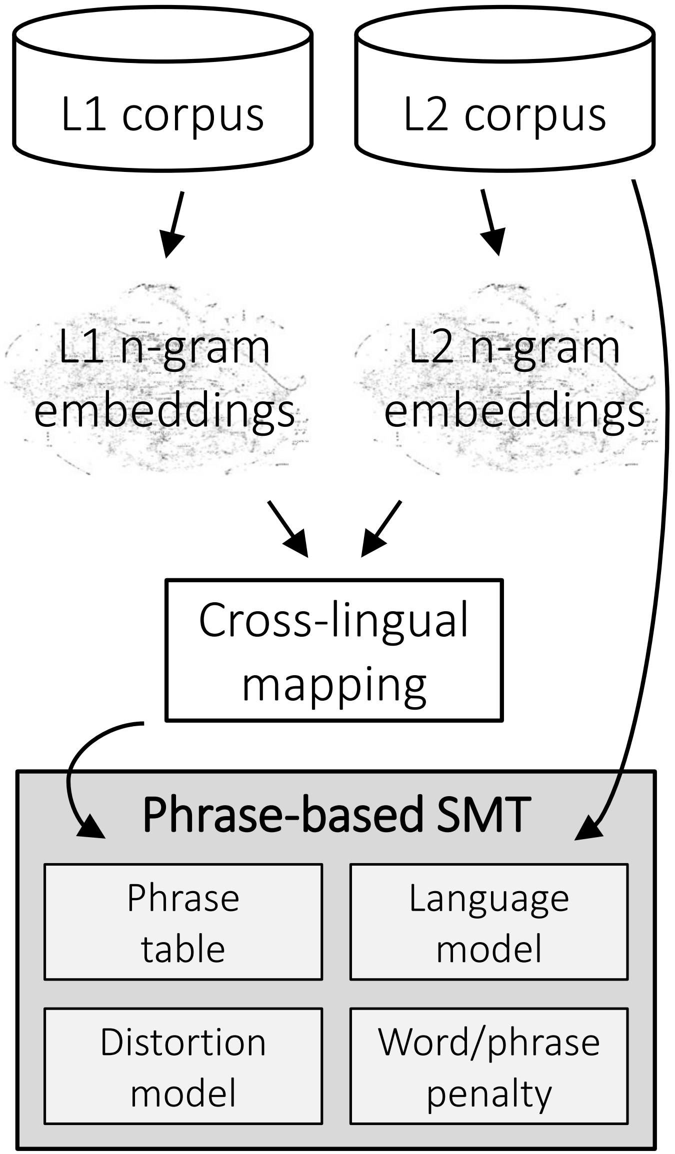
The basic approach takes words as atomic units



Unsupervised phrase-based SMT

The basic approach takes words as atomic units

Difficulties to translate named entities

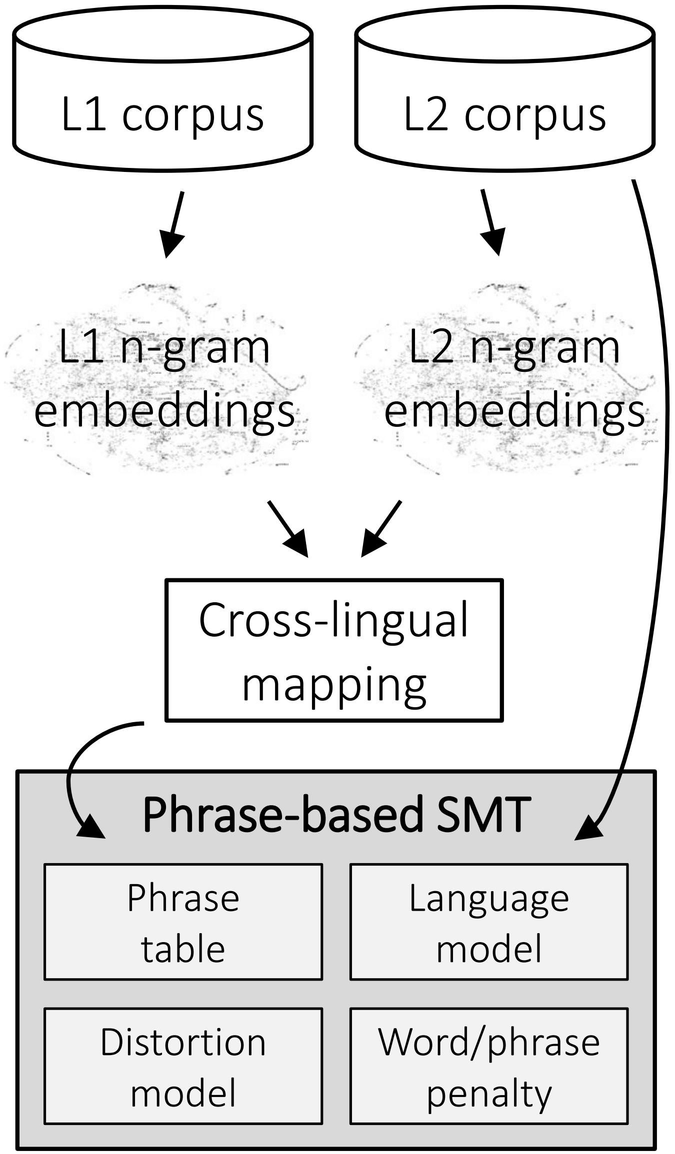


Unsupervised phrase-based SMT

The basic approach takes words as atomic units

Difficulties to translate named entities

- *“Sunday Telegraph”* → *“The Times of London”* (Artetxe et al., EMNLP’18)



Unsupervised phrase-based SMT

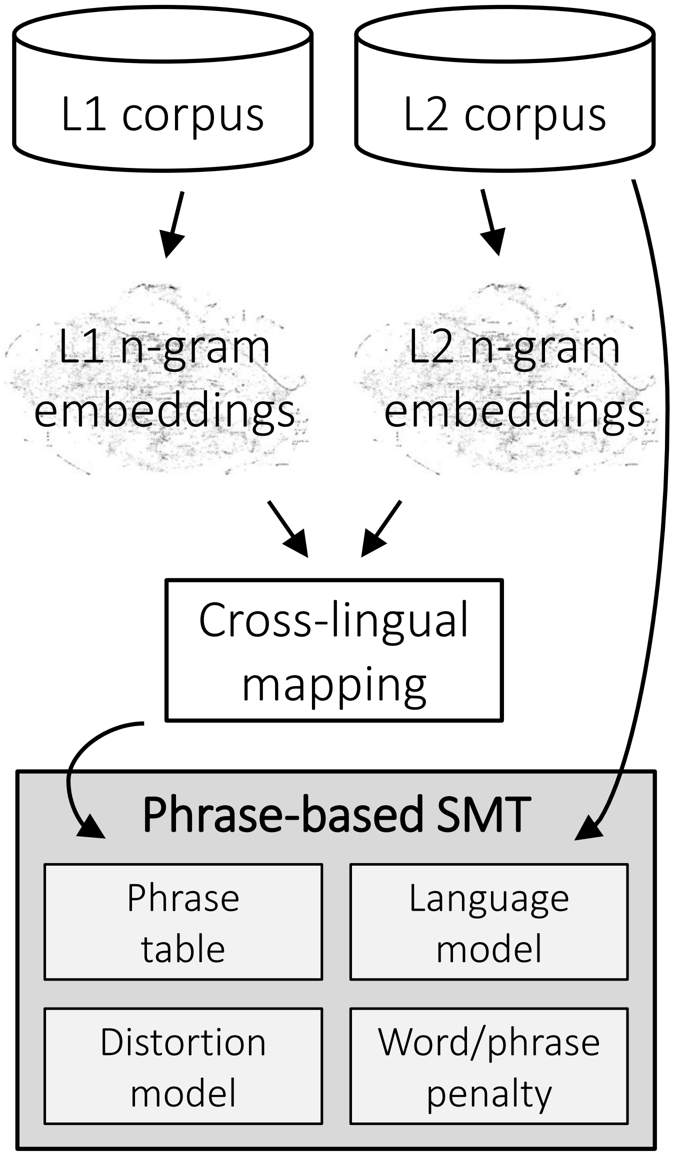
The basic approach takes words as atomic units

Difficulties to translate named entities

- *"Sunday Telegraph"* → *"The Times of London"* (Artetxe et al., EMNLP'18)

Add new features into the phrase-table

Unsupervised phrase-based SMT



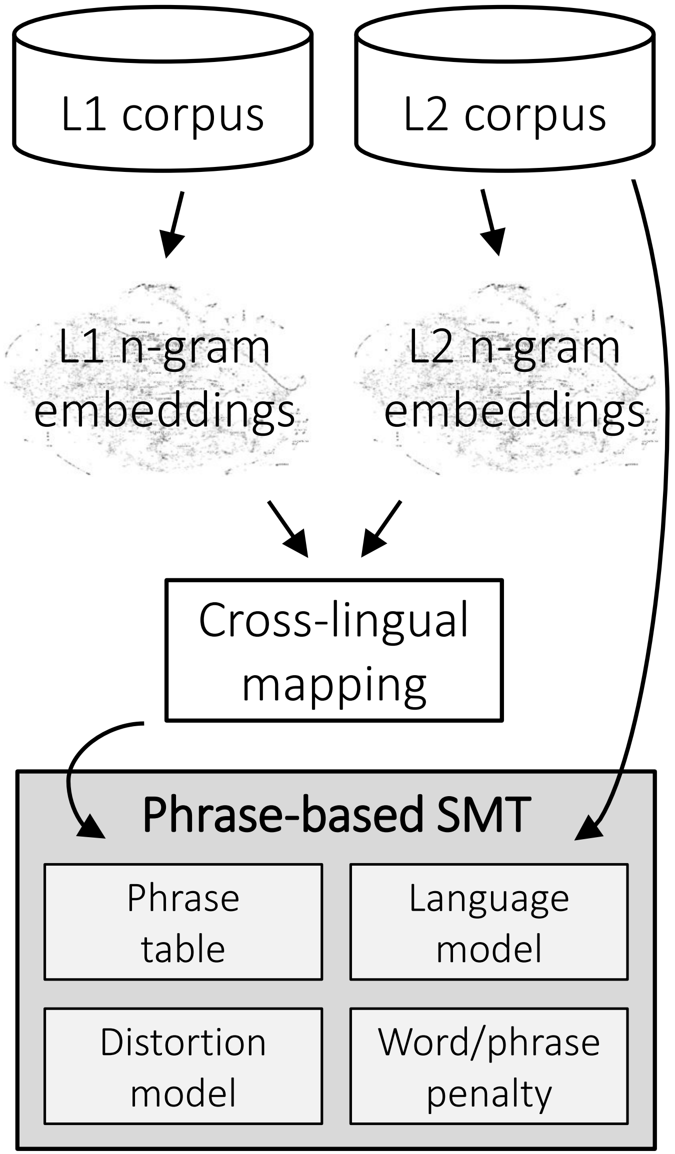
The basic approach takes words as atomic units

Difficulties to translate named entities

- *"Sunday Telegraph"* → *"The Times of London"* (Artetxe et al., EMNLP'18)

Add new features into the phrase-table

$$\text{score}(\bar{f}|\bar{e}) = \prod_i \max\left(\epsilon, \max_j \text{sim}(\bar{f}_i, \bar{e}_j)\right)$$



Unsupervised phrase-based SMT

The basic approach takes words as atomic units

Difficulties to translate named entities

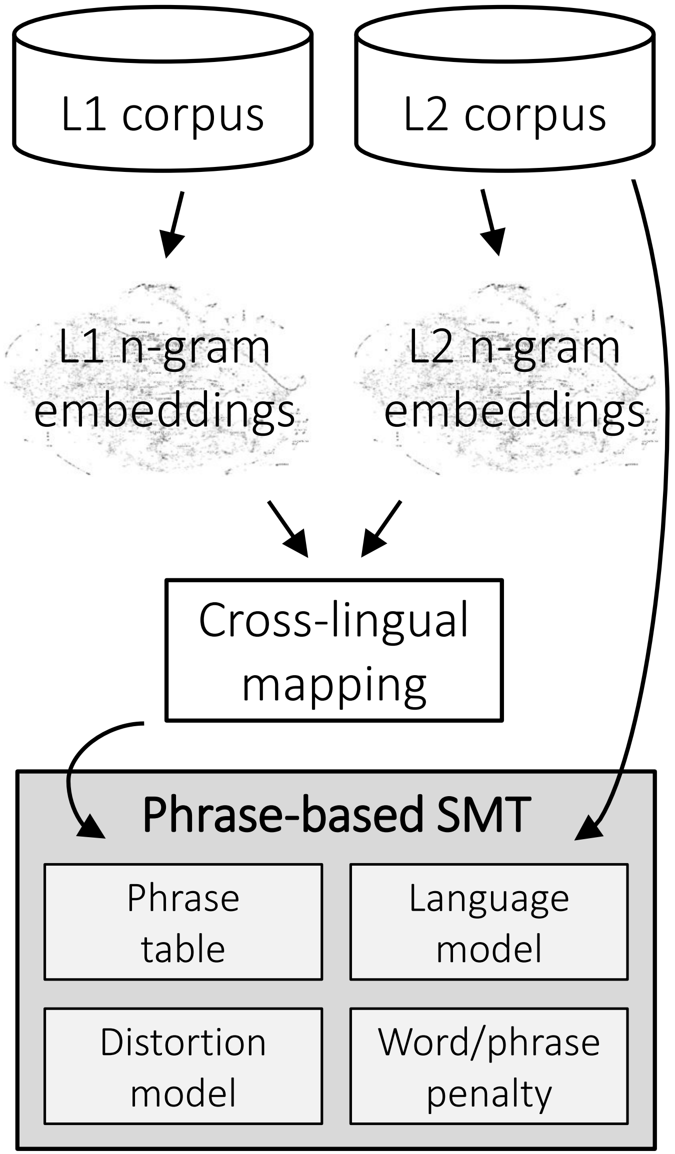
- *"Sunday Telegraph"* → *"The Times of London"* (Artetxe et al., EMNLP'18)

Add new features into the phrase-table

$$\mathrm{score}(\bar{f}|\bar{e}) = \prod_i \max\left(\epsilon, \max_j \mathrm{sim}(\bar{f}_i, \bar{e}_j)\right)$$

$$\mathrm{sim}(f, e) = 1 - \frac{\mathrm{lev}(f, e)}{\max(\mathrm{len}(f), \mathrm{len}(e))}$$

Unsupervised phrase-based SMT

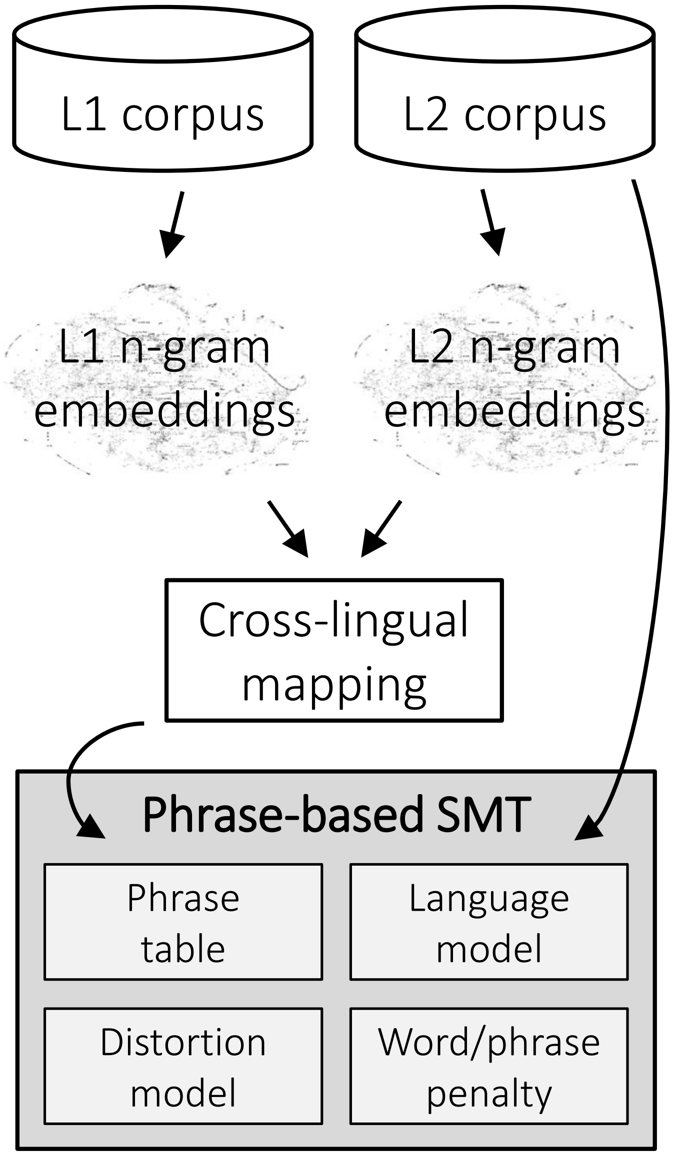


EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

	FR-EN	EN-FR	DE-EN	EN-DE
NMT (ICLR'18)*	15.6	15.1	10.2	6.6

*Tokenized BLEU (about 1-2 points higher)



Unsupervised phrase-based SMT

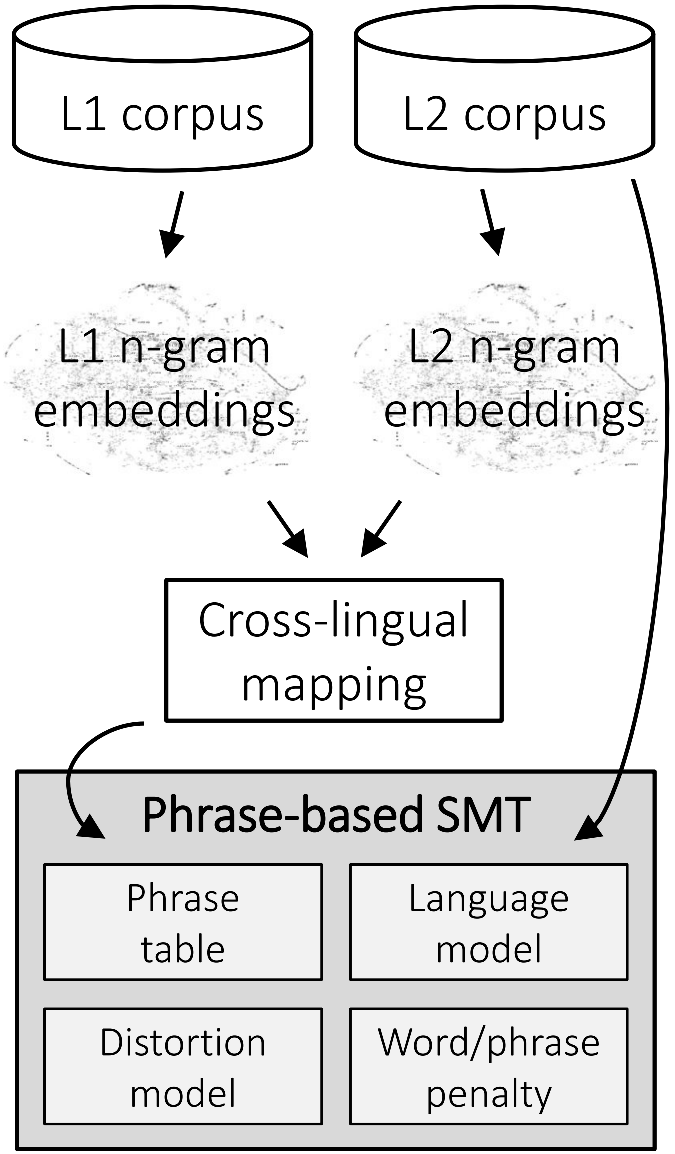
EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

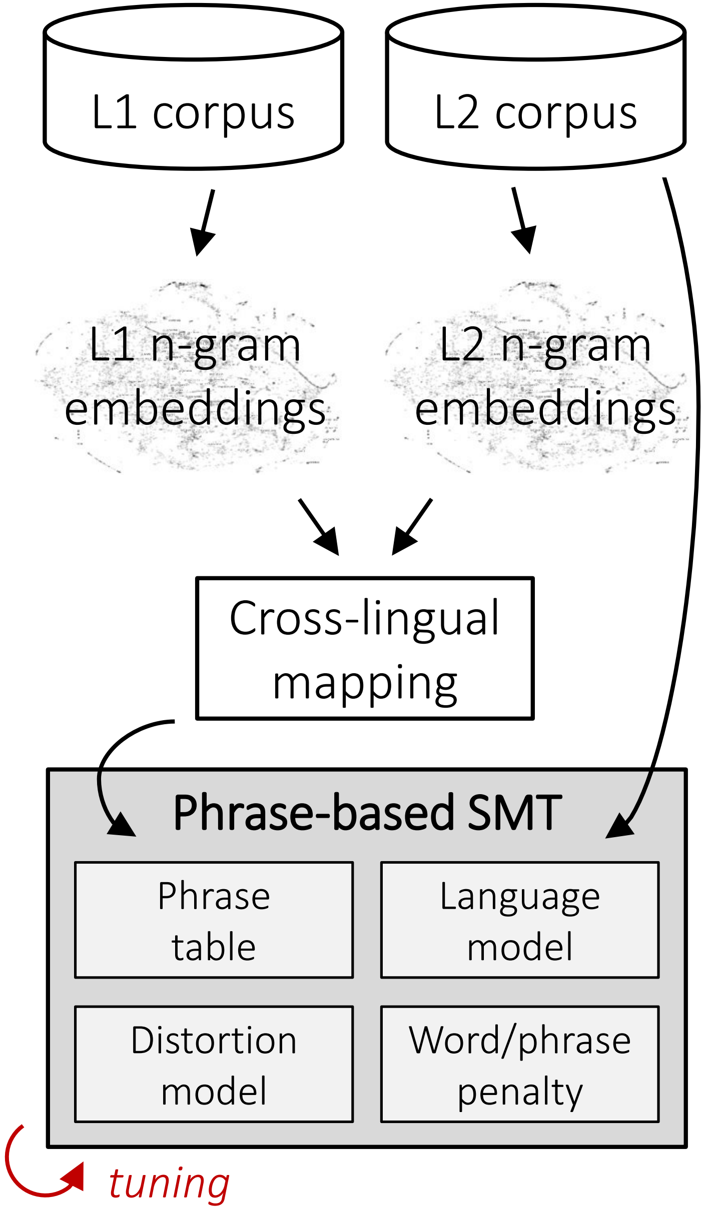
	FR-EN	EN-FR	DE-EN	EN-DE
NMT (ICLR'18)*	15.6	15.1	10.2	6.6
Initial SMT (ACL'19)	22.4	19.6	15.3	11.0

*Tokenized BLEU (about 1-2 points higher)

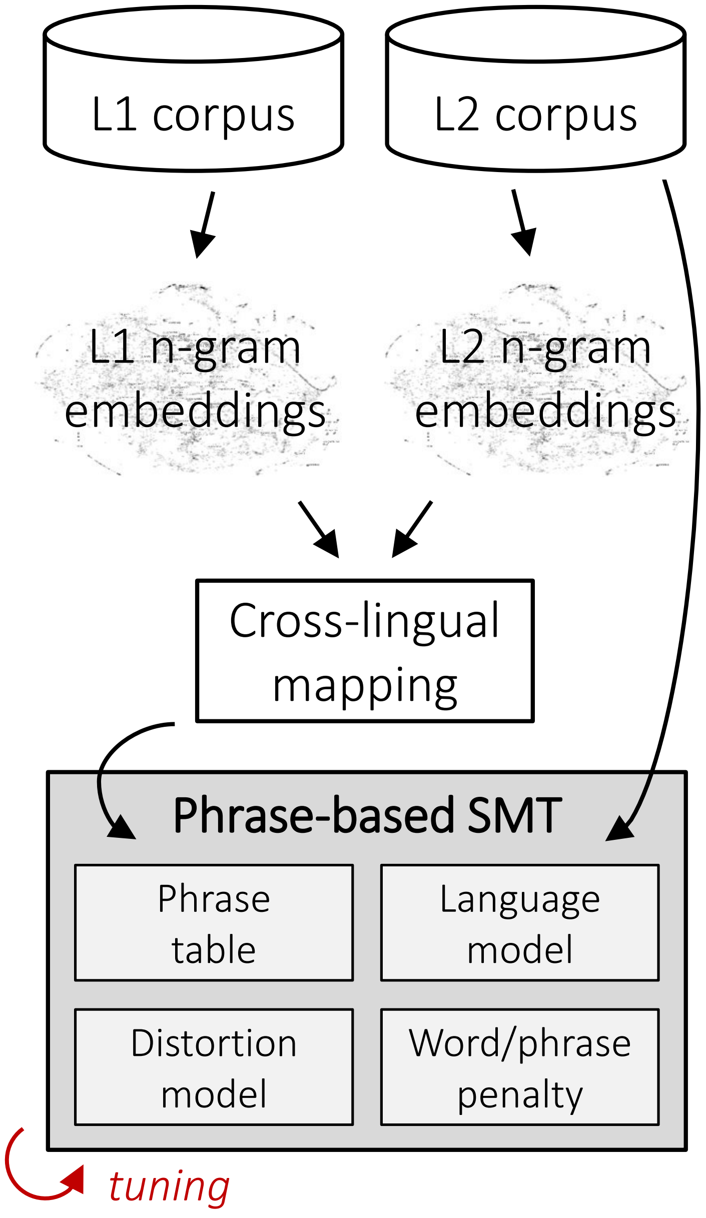
Unsupervised phrase-based SMT



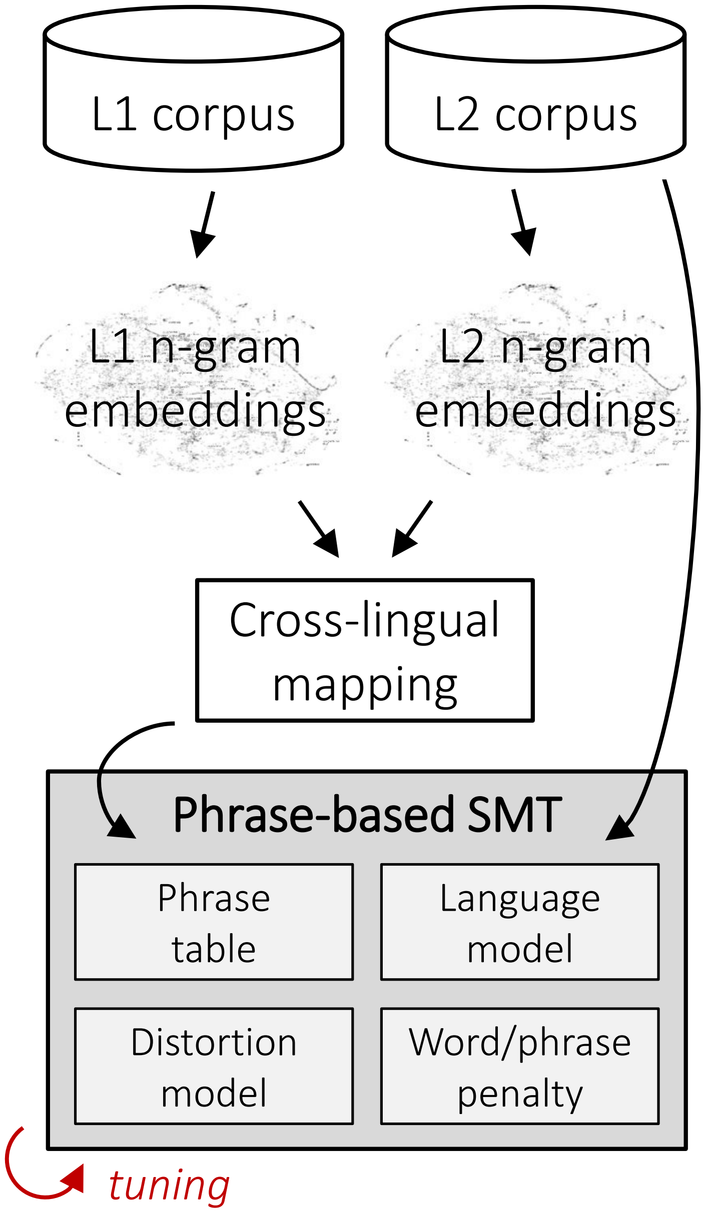
Unsupervised phrase-based SMT



Tuning

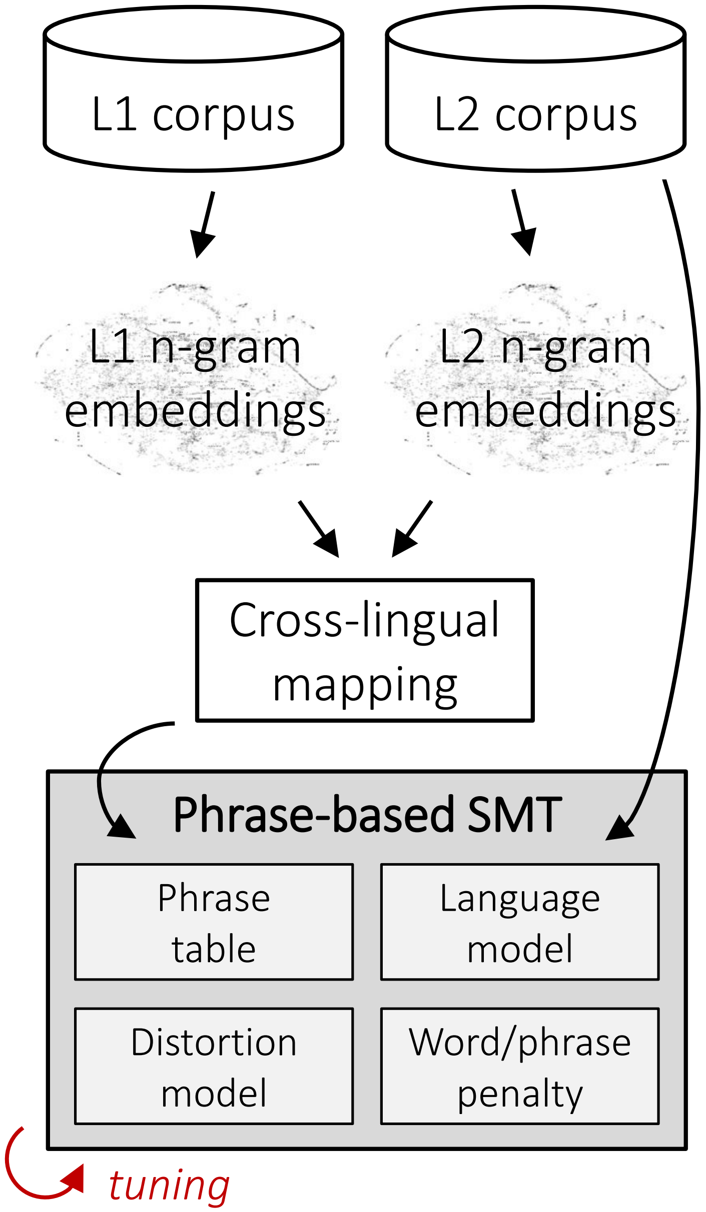


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning



Tuning

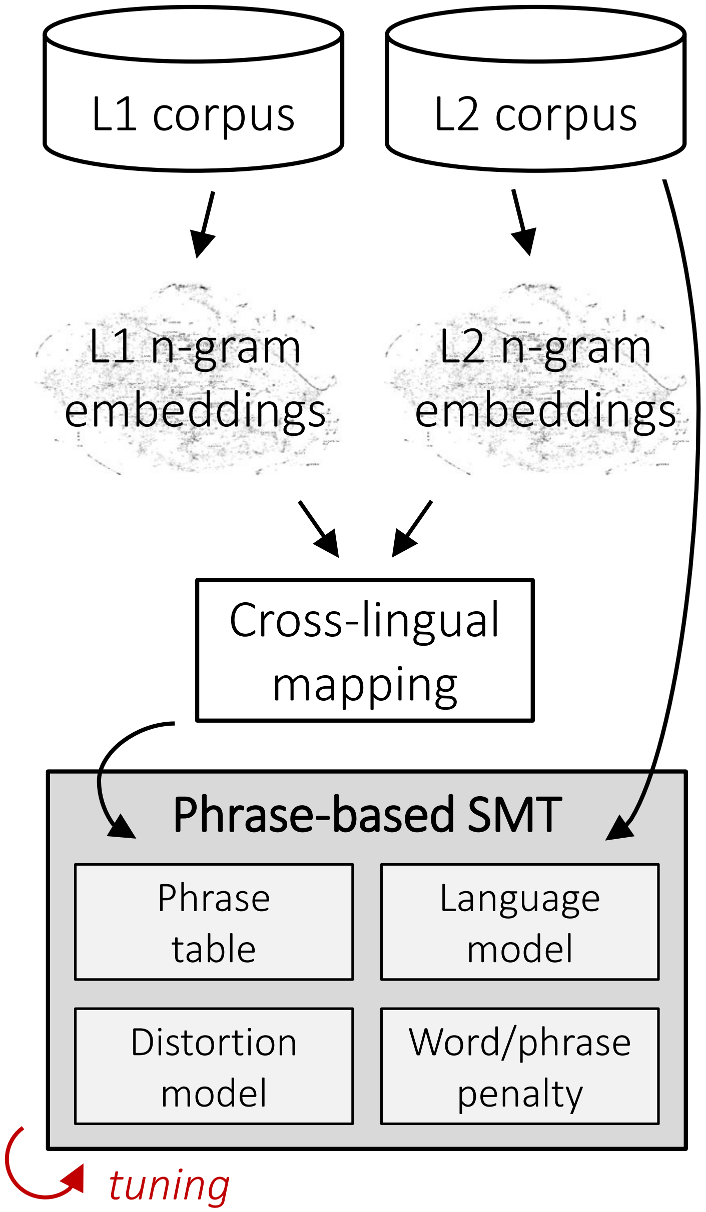
Goal: Adjust the weights of the resulting log-linear model to optimize some evaluation metric (e.g. BLEU)



Tuning

Goal: Adjust the weights of the resulting log-linear model to optimize some evaluation metric (e.g. BLEU)

… but we don't have a parallel development set!

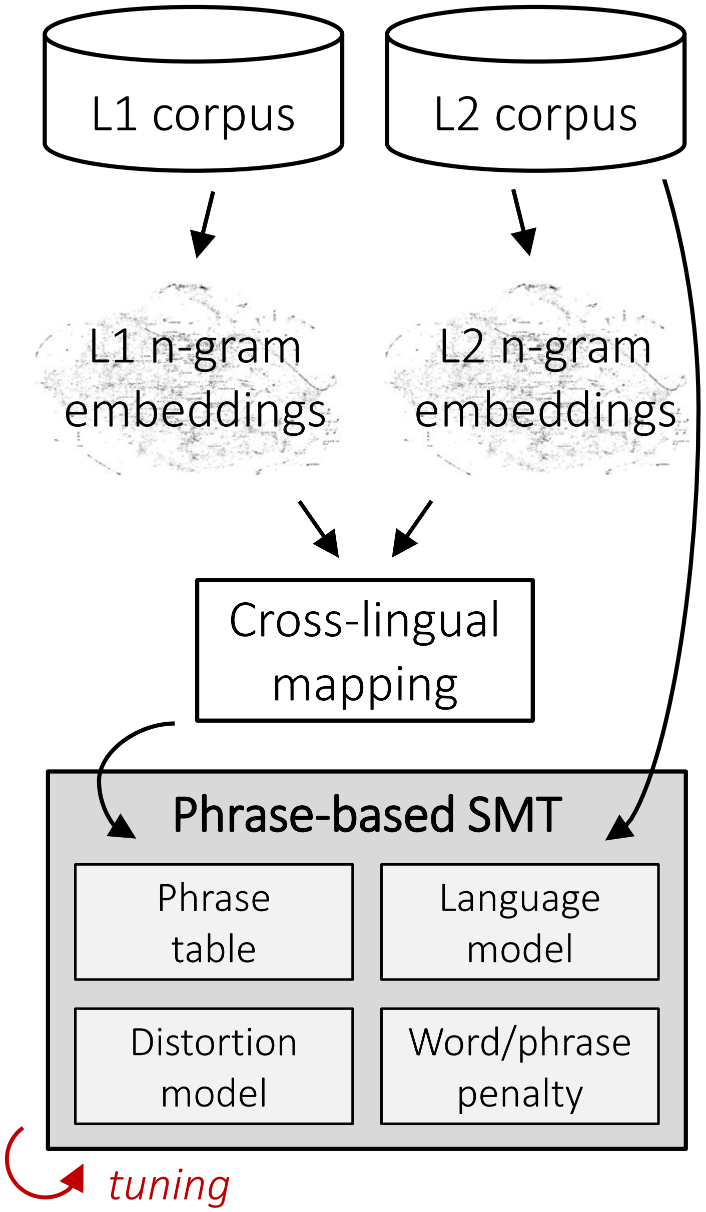


Tuning

Goal: Adjust the weights of the resulting log-linear model to optimize some evaluation metric (e.g. BLEU)

... but we don't have a parallel development set!

Early work



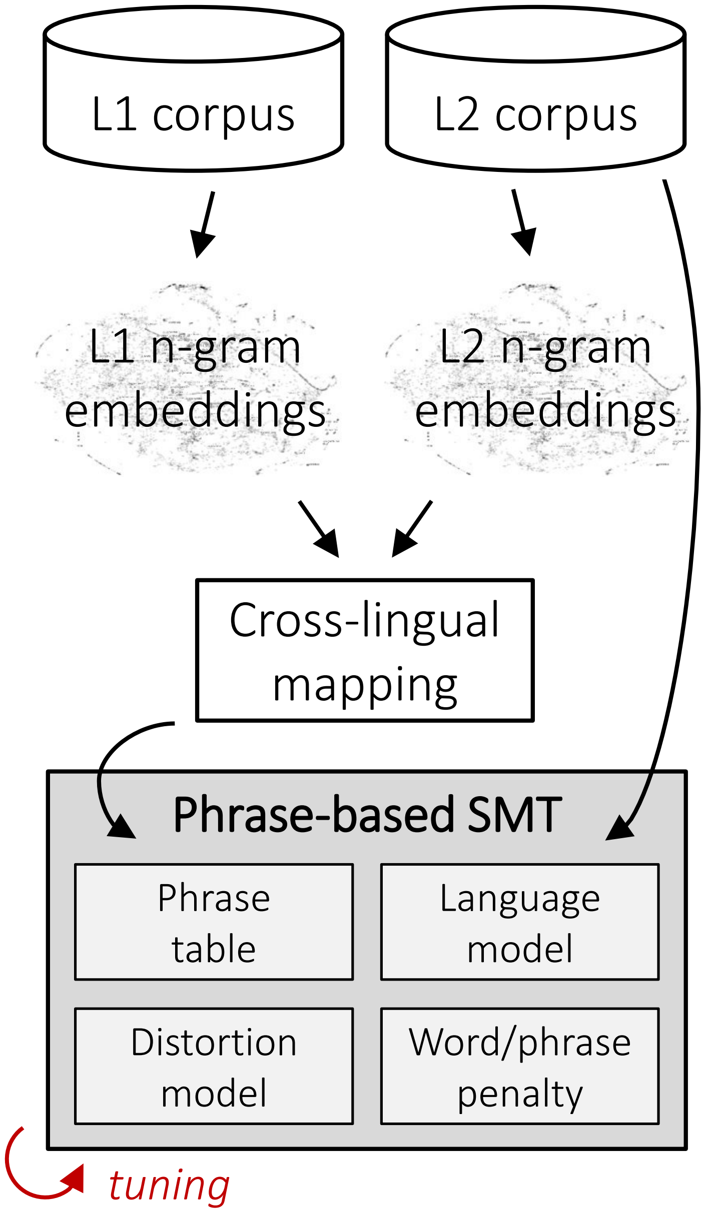
Tuning

Goal: Adjust the weights of the resulting log-linear model to optimize some evaluation metric (e.g. BLEU)

… but we don't have a parallel development set!

Early work

- Default weights without tuning (Lample et al., EMNLP'18)



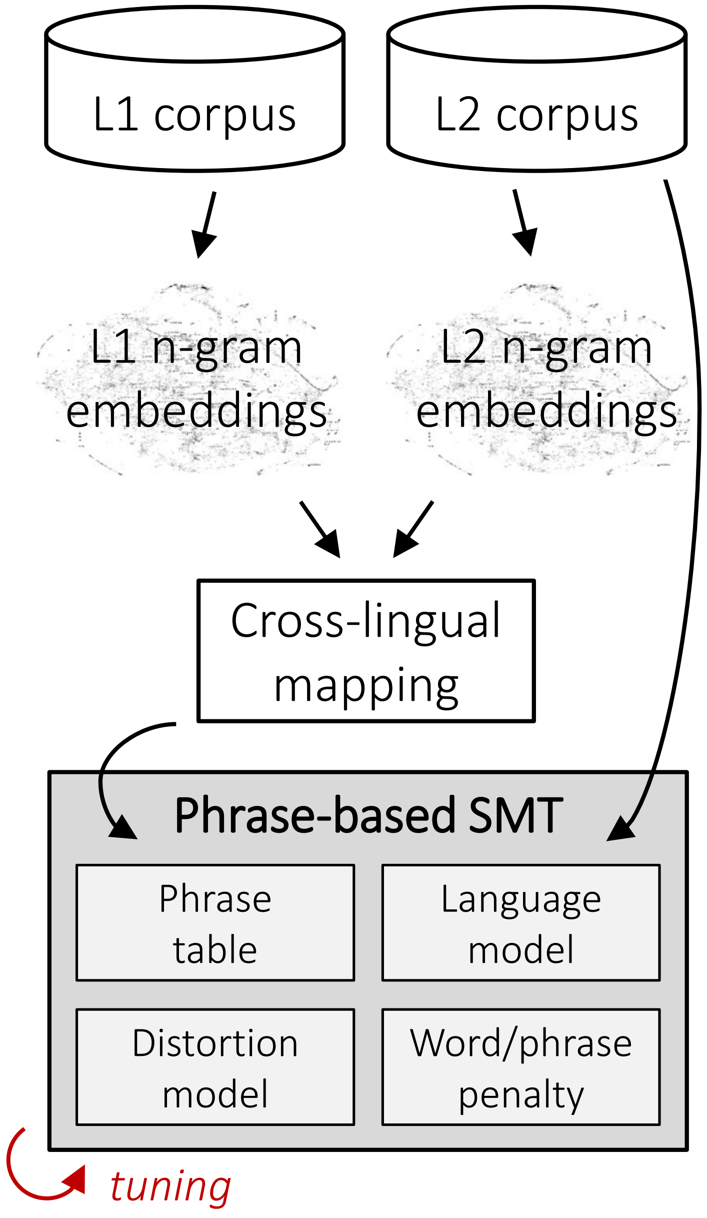
Tuning

Goal: Adjust the weights of the resulting log-linear model to optimize some evaluation metric (e.g. BLEU)

… but we don't have a parallel development set!

Early work

- Default weights without tuning (Lample et al., EMNLP'18)
- Alternating optimization with back-translation (Artetxe et al., EMNLP'18)



Tuning

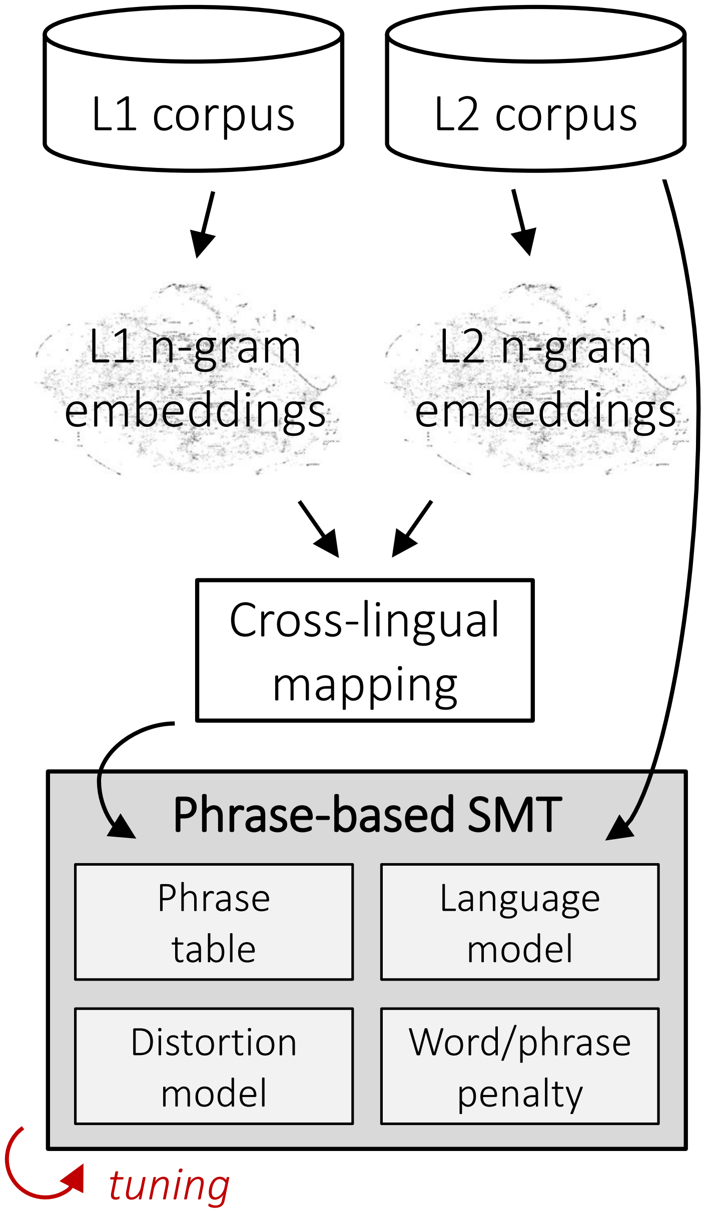
Goal: Adjust the weights of the resulting log-linear model to optimize some evaluation metric (e.g. BLEU)

… but we don't have a parallel development set!

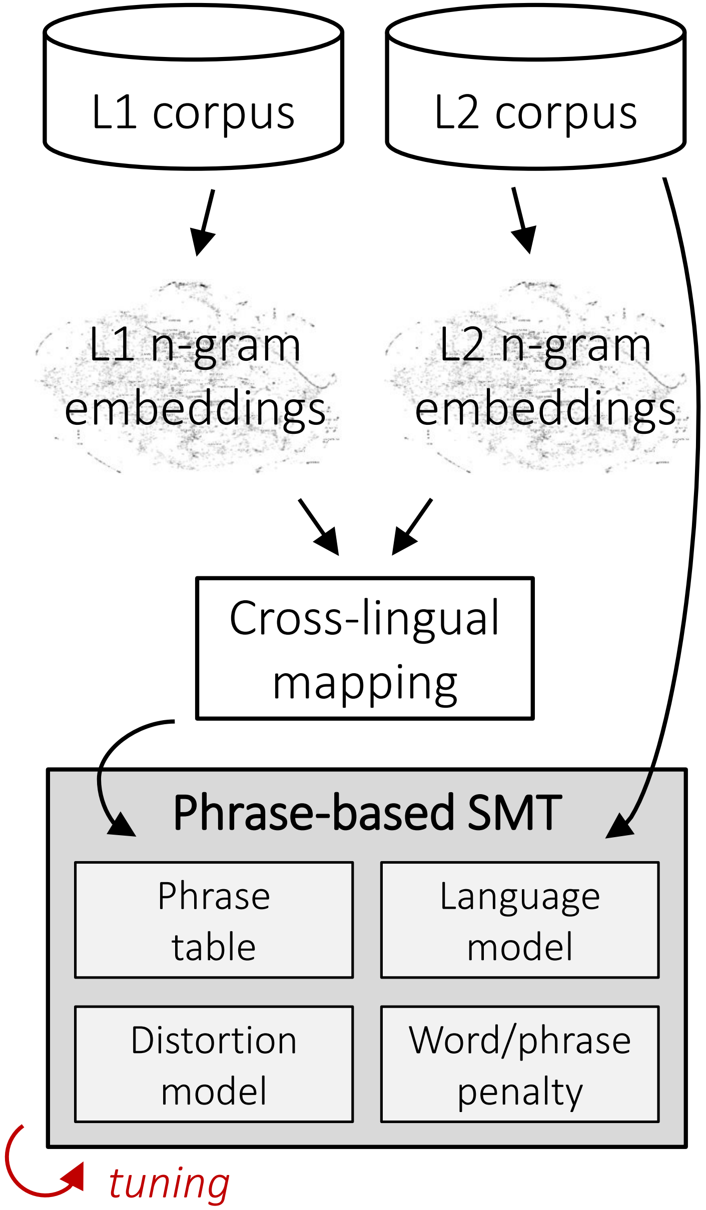
Early work

- Default weights without tuning (Lample et al., EMNLP'18)
- Alternating optimization with back-translation (Artetxe et al., EMNLP'18)

…let's build a more principled approach!

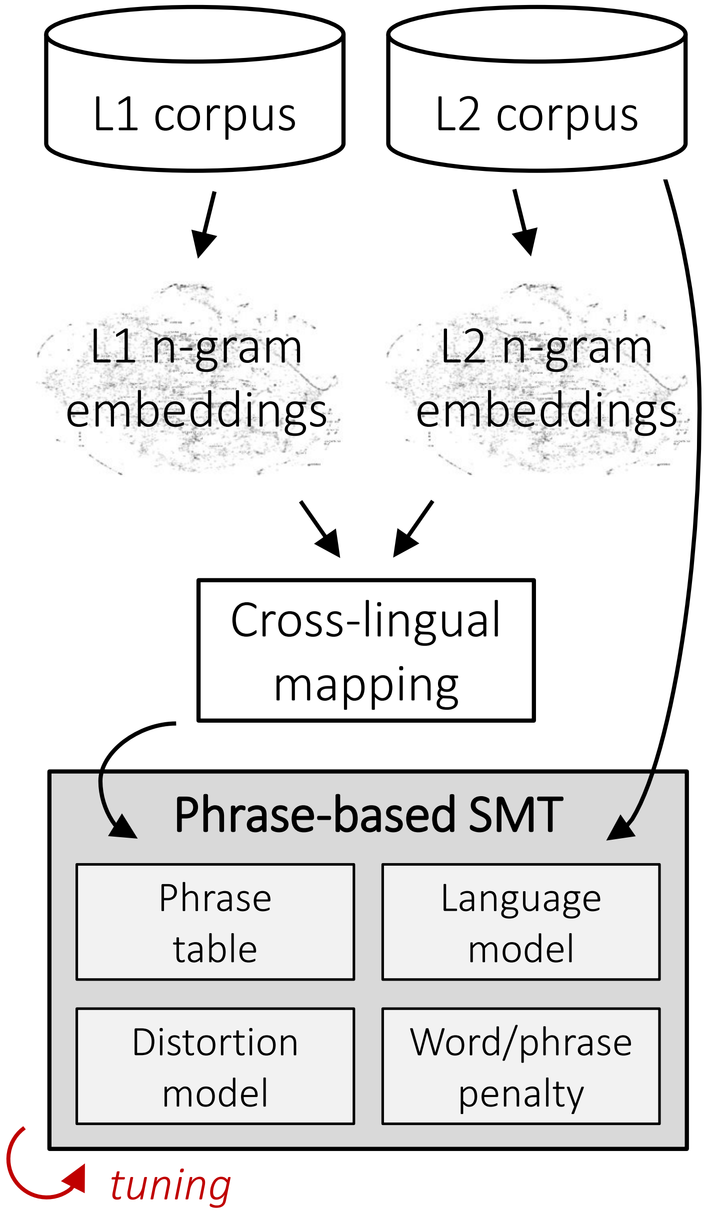

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning

Tuning



Tuning

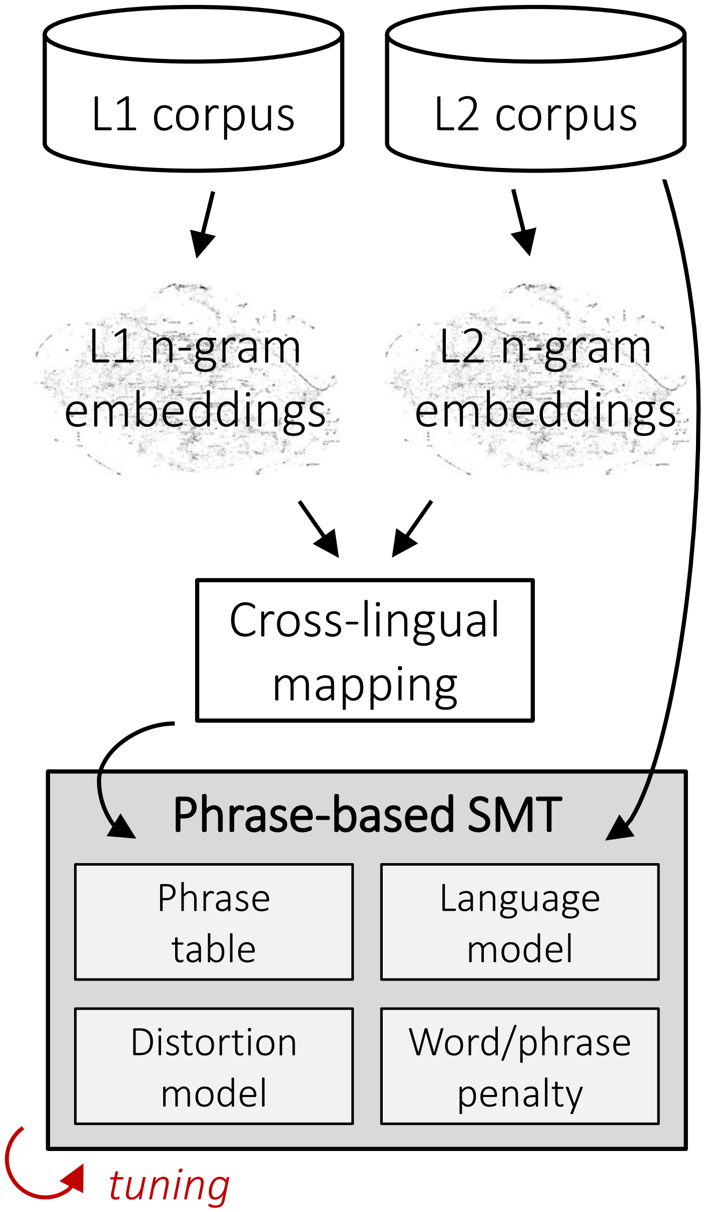
Unsupervised optimization objective



Tuning

Unsupervised optimization objective

$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

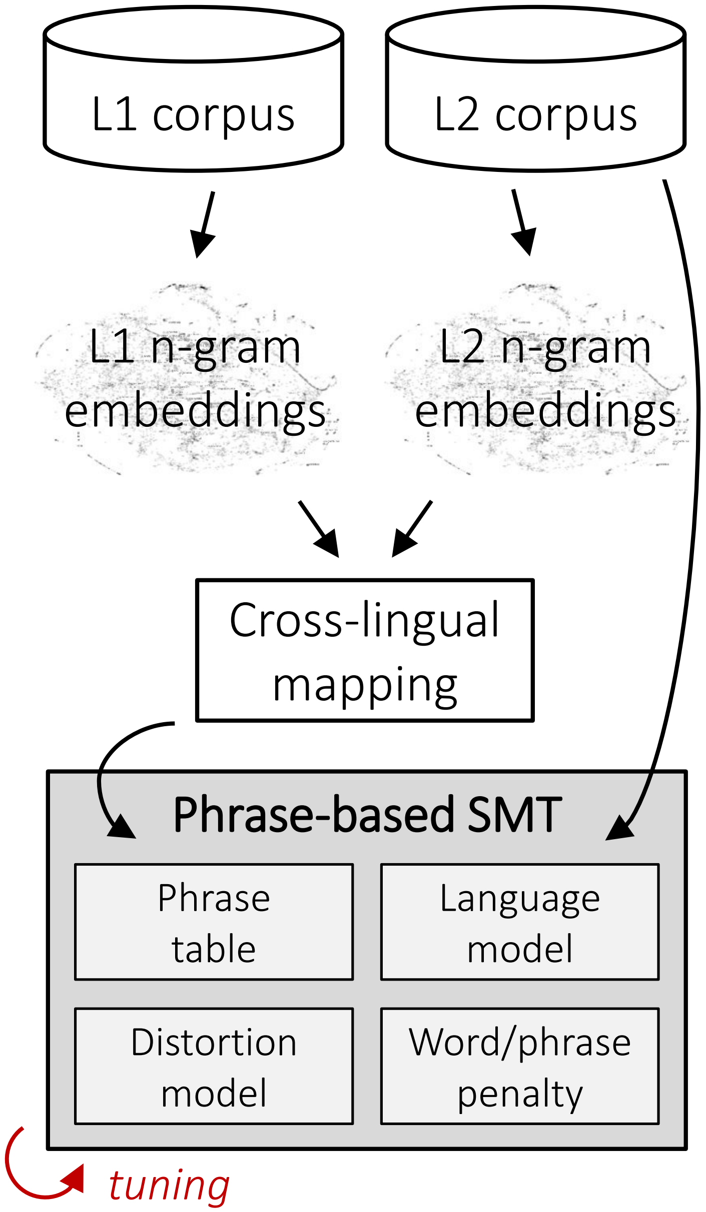


Tuning

Unsupervised optimization objective

$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

- $L_{cycle}(E) = 1 - \mathrm{BLEU}\big(\mathrm{T}_{F\to E}\big(\mathrm{T}_{E\to F}(E)\big), E\big)$

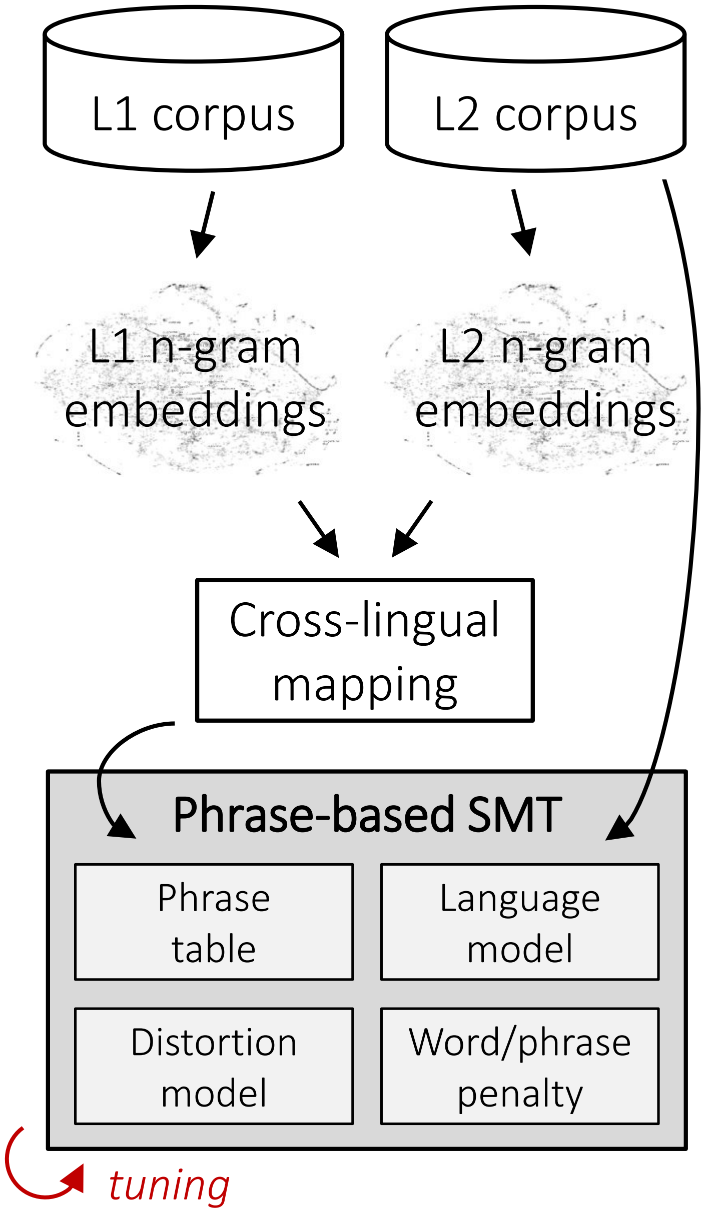


Tuning

Unsupervised optimization objective

$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

- $L_{cycle}(E) = 1 - \mathrm{BLEU}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right), E\right)$
- $L_{lm}(E) = \max\left(0, \mathrm{H}(F) - \mathrm{H}\left(\mathrm{T}_{E\to F}(E)\right)\right)^2$

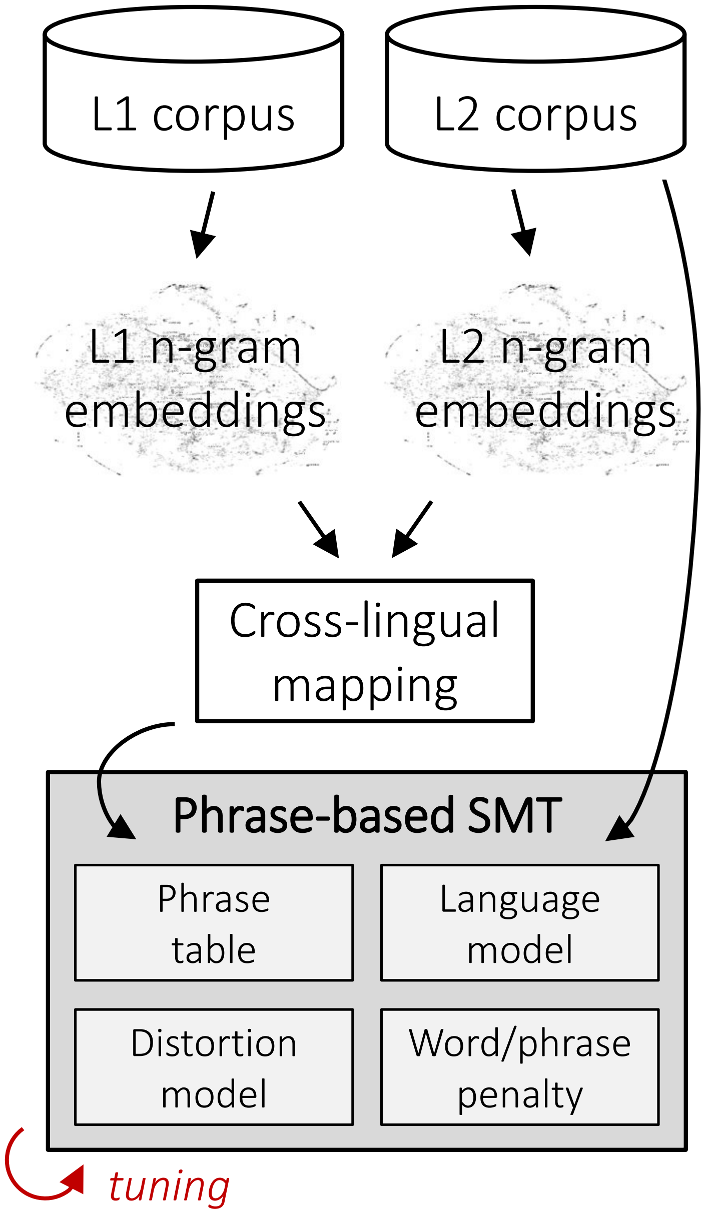


Tuning

Unsupervised optimization objective

$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

- $L_{cycle}(E) = 1 - \mathrm{BLEU}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right), E\right)$
- $L_{lm}(E) = \max\left(0, \mathrm{H}(F) - \mathrm{H}\left(\mathrm{T}_{E\to F}(E)\right)\right)^2 \cdot \mathrm{LP}$



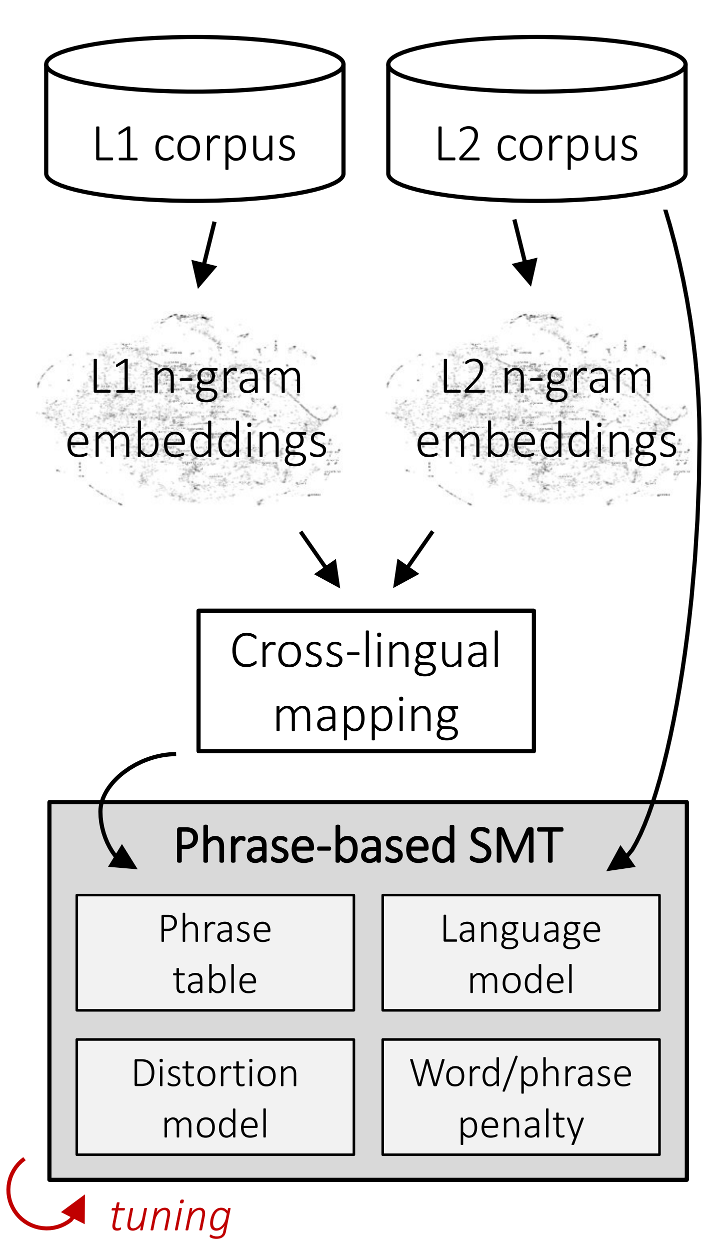
Tuning

Unsupervised optimization objective

$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

- $L_{cycle}(E) = 1 - \mathrm{BLEU}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right), E\right)$
- $L_{lm}(E) = \max\left(0, \mathrm{H}(F) - \mathrm{H}\left(\mathrm{T}_{E\to F}(E)\right)\right)^2 \cdot \mathrm{LP}$

$$\mathrm{LP} = \mathrm{LP}(E) \cdot \mathrm{LP}(F)\,, \quad \mathrm{LP}(E) = \max\left(1, \frac{\mathrm{len}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right)\right)}{\mathrm{len}(E)}\right)$$



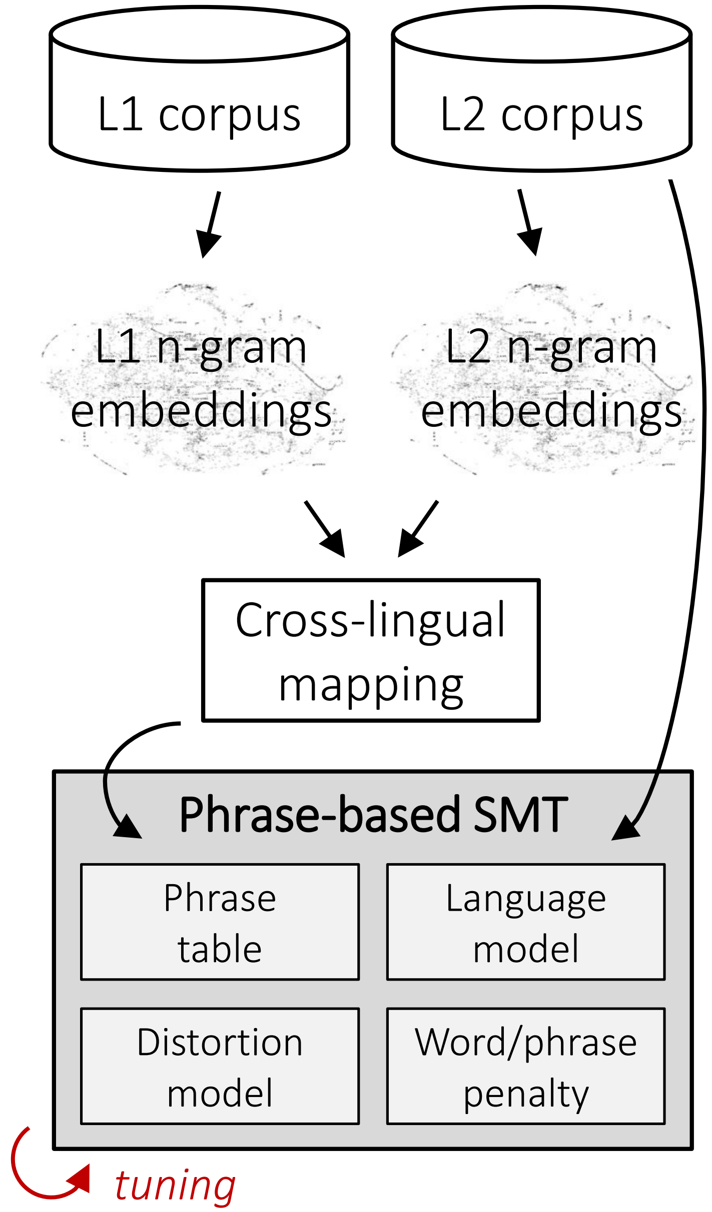
Tuning

Unsupervised optimization objective

$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

- $L_{cycle}(E) = 1 - \mathrm{BLEU}\big(\mathrm{T}_{F\to E}\big(\mathrm{T}_{E\to F}(E)\big), E\big)$
- $L_{lm}(E) = \max\Big(0, \mathrm{H}(F) - \mathrm{H}\big(\mathrm{T}_{E\to F}(E)\big)\Big)^2 \cdot \mathrm{LP}$

$$\mathrm{LP} = \mathrm{LP}(E) \cdot \mathrm{LP}(F)\,, \quad \mathrm{LP}(E) = \max\left(1, \frac{\mathrm{len}\big(\mathrm{T}_{F\to E}(\mathrm{T}_{E\to F}(E))\big)}{\mathrm{len}(E)}\right)$$



Tuning

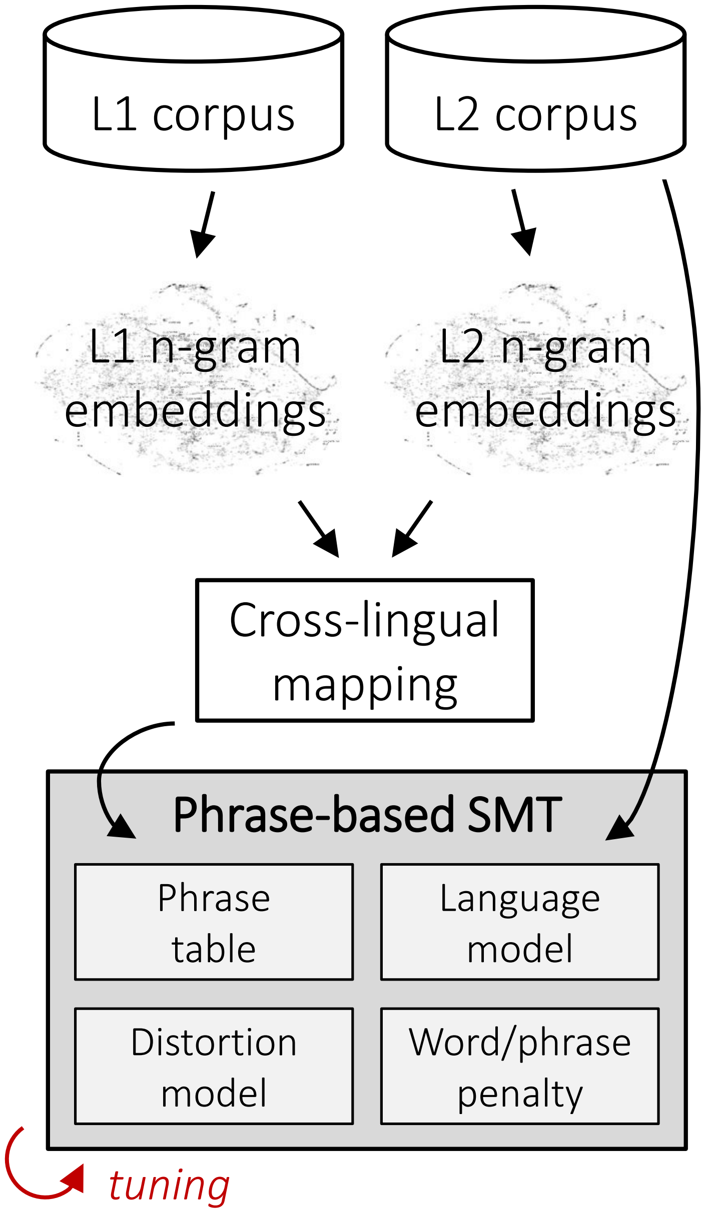
Unsupervised optimization objective

$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

- $L_{cycle}(E) = 1 - \mathrm{BLEU}\left(\boxed{\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}\right.}(E)\right), E\Big)$
- $L_{lm}(E) = \max\left(0, \mathrm{H}(F) - \mathrm{H}\left(\mathrm{T}_{E\to F}(E)\right)\right)^2 \cdot \mathrm{LP}$

$$\mathrm{LP} = \mathrm{LP}(E) \cdot \mathrm{LP}(F)\,, \quad \mathrm{LP}(E) = \max\left(1, \frac{\mathrm{len}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right)\right)}{\mathrm{len}(E)}\right)$$

Regular MERT would require a combined n-best list of n^2 entries!



Tuning

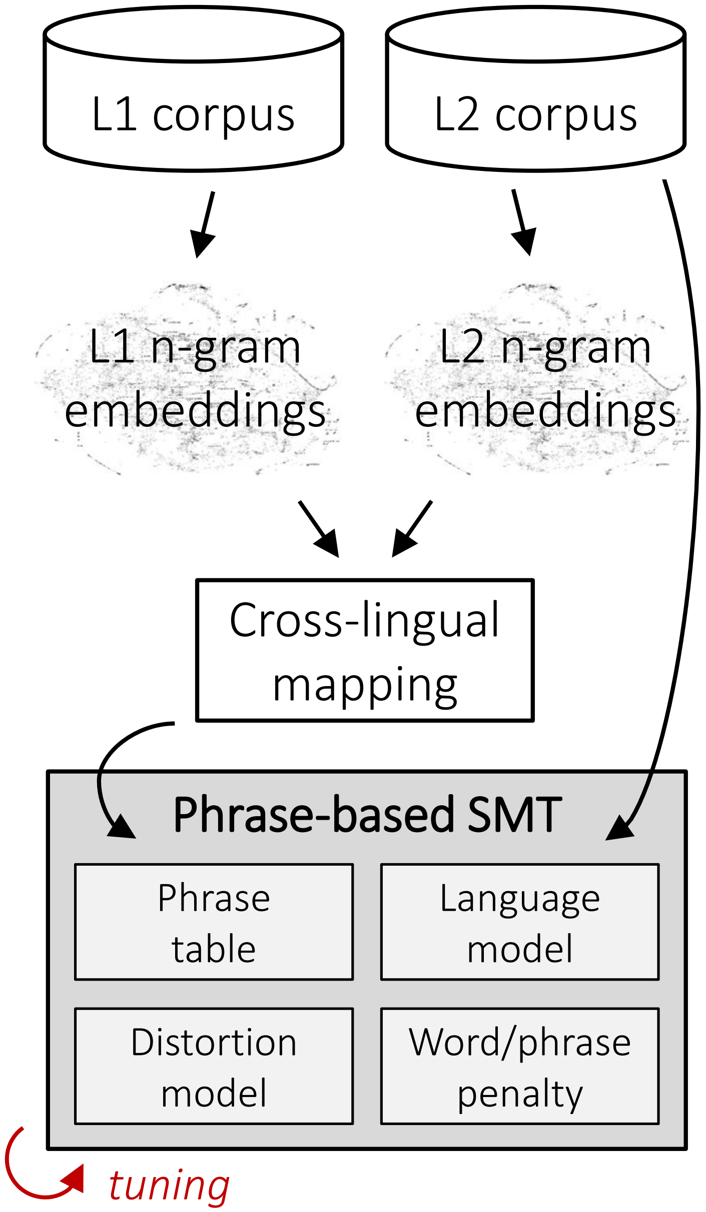
Unsupervised optimization objective

$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

- $L_{cycle}(E) = 1 - \mathrm{BLEU}\big(\mathrm{T}_{F\to E}\big(\mathrm{T}_{E\to F}(E)\big), E\big)$
- $L_{lm}(E) = \max\Big(0, \mathrm{H}(F) - \mathrm{H}\big(\mathrm{T}_{E\to F}(E)\big)\Big)^2 \cdot \mathrm{LP}$

$$\mathrm{LP} = \mathrm{LP}(E)\cdot\mathrm{LP}(F)\,, \quad \mathrm{LP}(E) = \max\left(1, \frac{\mathrm{len}\big(\mathrm{T}_{F\to E}(\mathrm{T}_{E\to F}(E))\big)}{\mathrm{len}(E)}\right)$$

Alternating optimization



Tuning

Unsupervised optimization objective

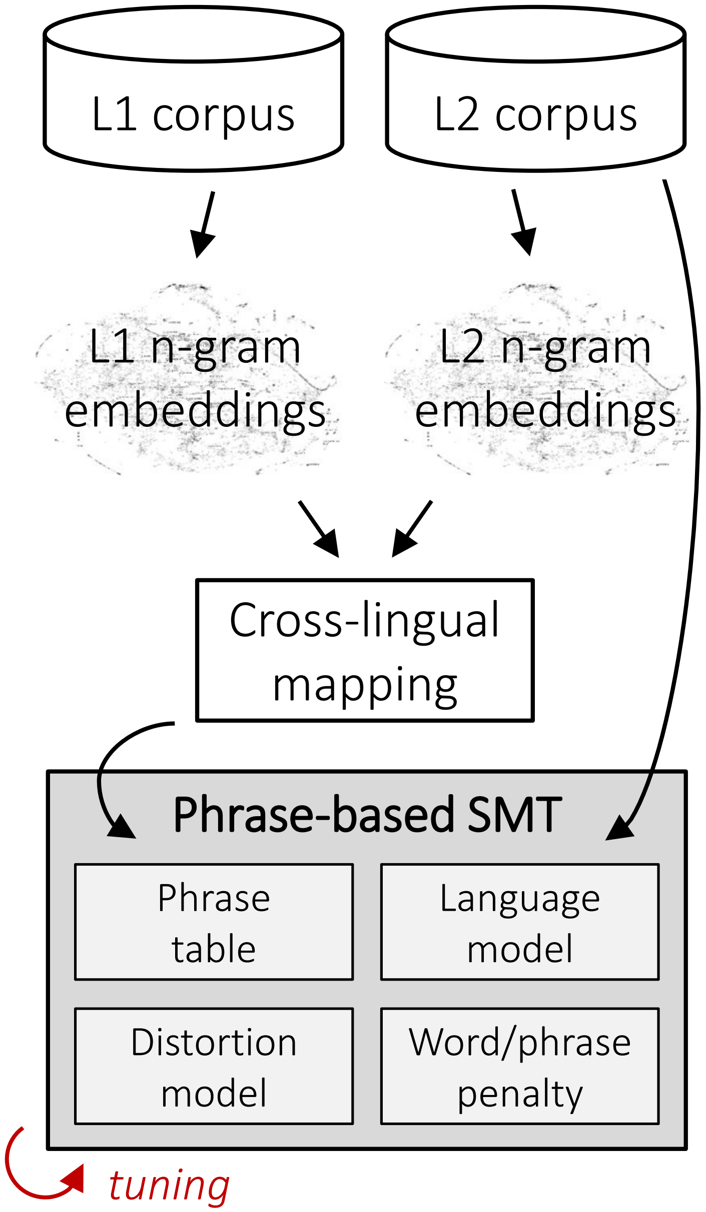
$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

- $L_{cycle}(E) = 1 - \mathrm{BLEU}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right), E\right)$
- $L_{lm}(E) = \max\left(0, \mathrm{H}(F) - \mathrm{H}\left(\mathrm{T}_{E\to F}(E)\right)\right)^2 \cdot \mathrm{LP}$

$$\mathrm{LP} = \mathrm{LP}(E) \cdot \mathrm{LP}(F)\,, \quad \mathrm{LP}(E) = \max\left(1, \frac{\mathrm{len}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right)\right)}{\mathrm{len}(E)}\right)$$

Alternating optimization

- Fix $\mathrm{T}_{F\to E}$ and optimize $\mathrm{T}_{E\to F}$ using MERT



Tuning

Unsupervised optimization objective

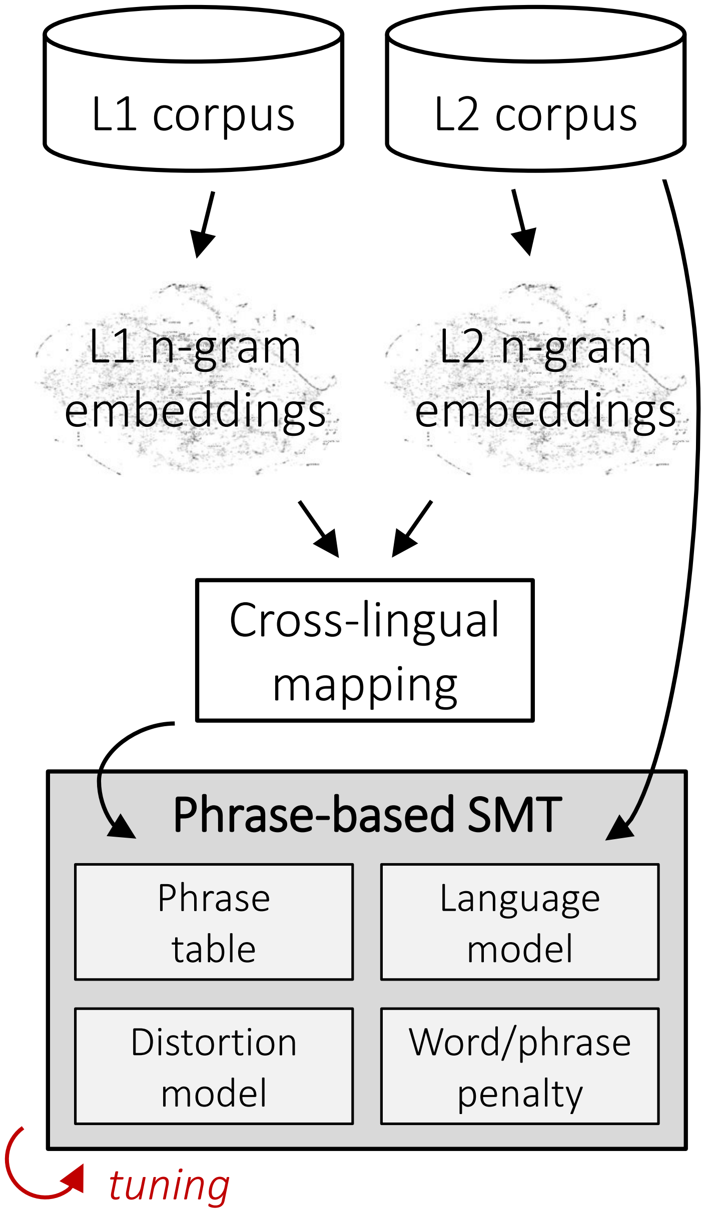
$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

- $L_{cycle}(E) = 1 - \mathrm{BLEU}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right), E\right)$
- $L_{lm}(E) = \max\left(0, \mathrm{H}(F) - \mathrm{H}\left(\mathrm{T}_{E\to F}(E)\right)\right)^2 \cdot \mathrm{LP}$

$$\mathrm{LP} = \mathrm{LP}(E) \cdot \mathrm{LP}(F)\,, \quad \mathrm{LP}(E) = \max\left(1, \frac{\mathrm{len}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right)\right)}{\mathrm{len}(E)}\right)$$

Alternating optimization

- Fix $\mathrm{T}_{F\to E}$ and optimize $\mathrm{T}_{E\to F}$ using MERT
- Fix $\mathrm{T}_{E\to F}$ and optimize $\mathrm{T}_{F\to E}$ using MERT



Tuning

Unsupervised optimization objective

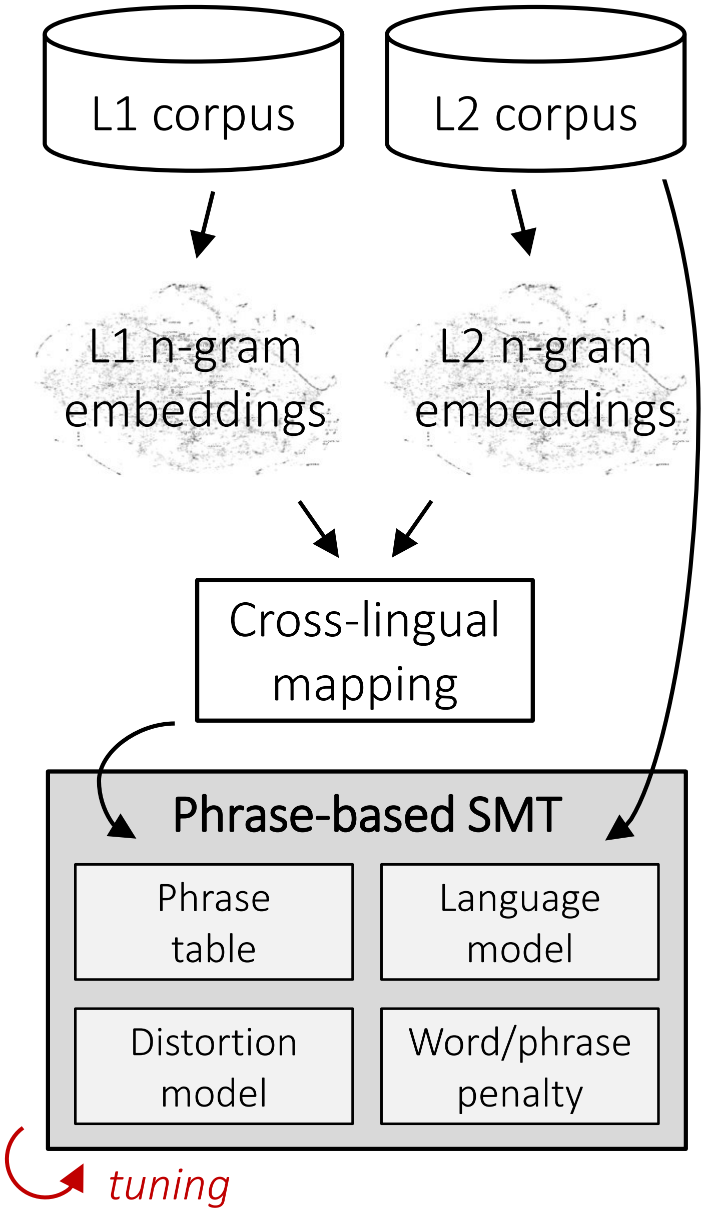
$$L = L_{cycle}(E) + L_{cycle}(F) + L_{lm}(E) + L_{lm}(F)$$

- $L_{cycle}(E) = 1 - \mathrm{BLEU}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right), E\right)$
- $L_{lm}(E) = \max\left(0, \mathrm{H}(F) - \mathrm{H}\left(\mathrm{T}_{E\to F}(E)\right)\right)^2 \cdot \mathrm{LP}$

$$\mathrm{LP} = \mathrm{LP}(E) \cdot \mathrm{LP}(F)\,, \quad \mathrm{LP}(E) = \max\left(1, \frac{\mathrm{len}\left(\mathrm{T}_{F\to E}\left(\mathrm{T}_{E\to F}(E)\right)\right)}{\mathrm{len}(E)}\right)$$

Alternating optimization

- Fix $\mathrm{T}_{F\to E}$ and optimize $\mathrm{T}_{E\to F}$ using MERT
- Fix $\mathrm{T}_{E\to F}$ and optimize $\mathrm{T}_{F\to E}$ using MERT
- Iterate until convergence



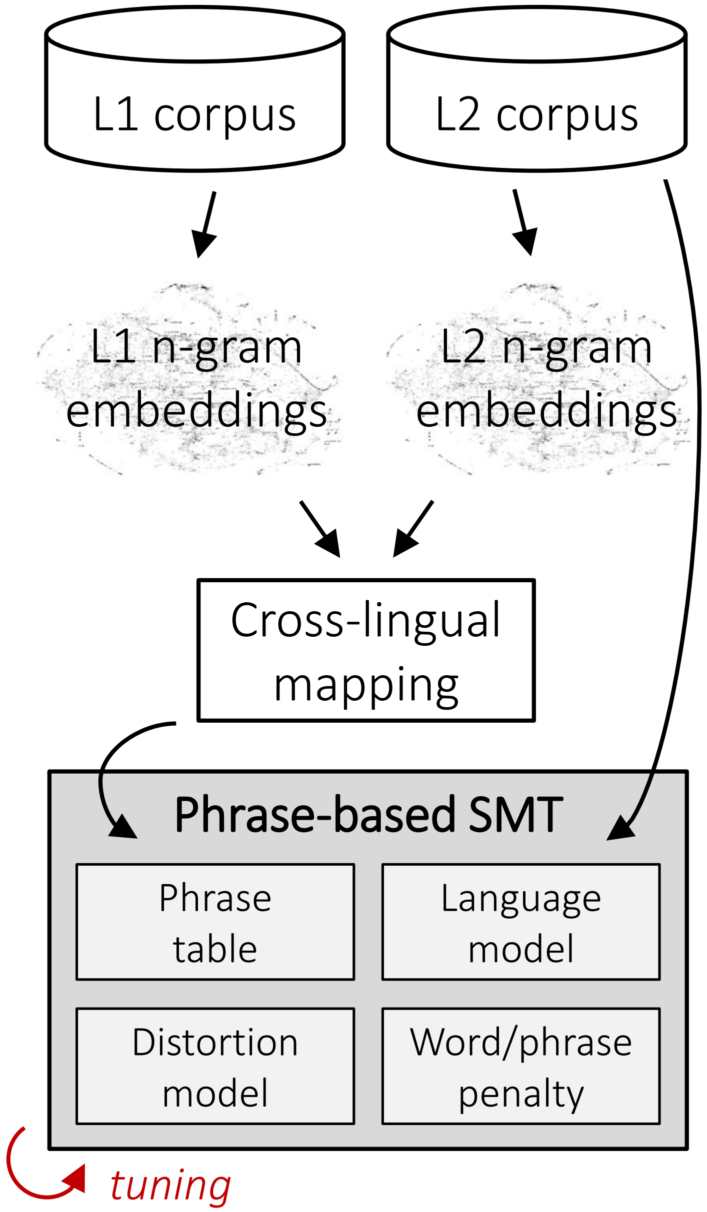
Tuning

EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

	FR-EN	EN-FR	DE-EN	EN-DE
NMT (ICLR'18)*	15.6	15.1	10.2	6.6
Initial SMT (ACL'19)	22.4	19.6	15.3	11.0

*Tokenized BLEU (about 1-2 points higher)



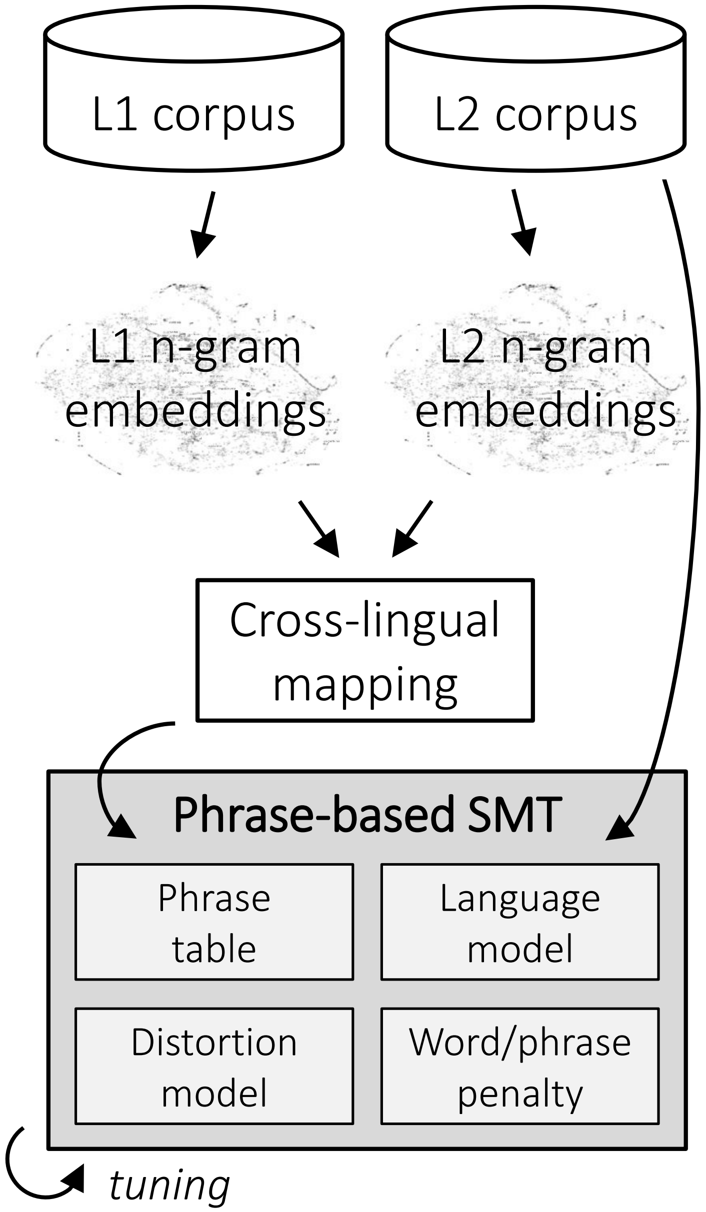
Tuning

EXPERIMENTS

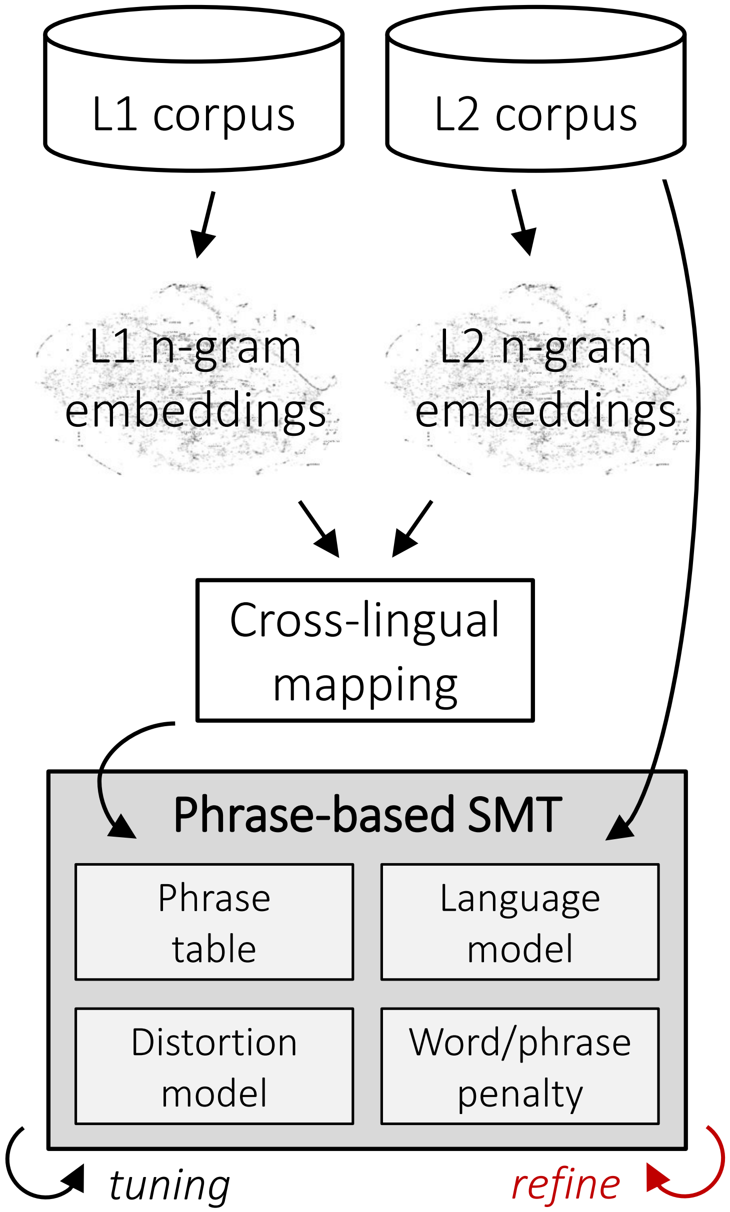
- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

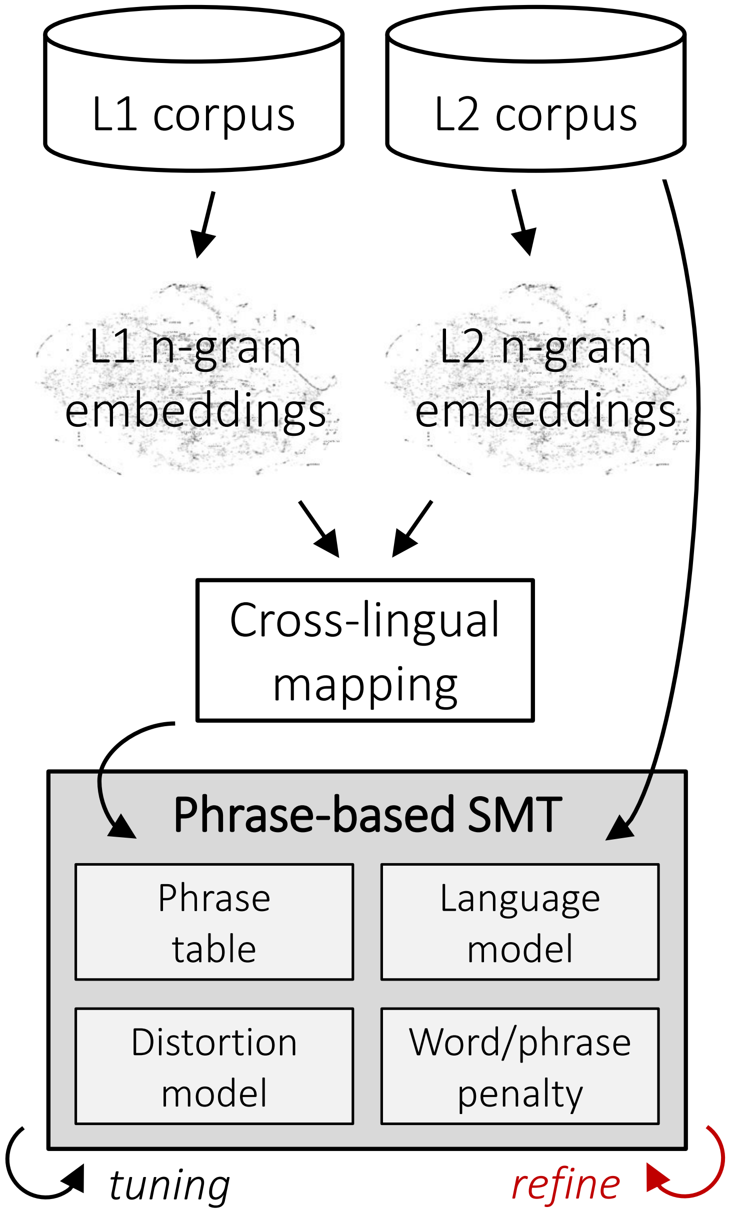
	FR-EN	EN-FR	DE-EN	EN-DE
NMT (ICLR'18)*	15.6	15.1	10.2	6.6
Initial SMT (ACL'19)	22.4	19.6	15.3	11.0
+ Tuning	23.4	21.9	15.4	11.2

*Tokenized BLEU (about 1-2 points higher)


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning

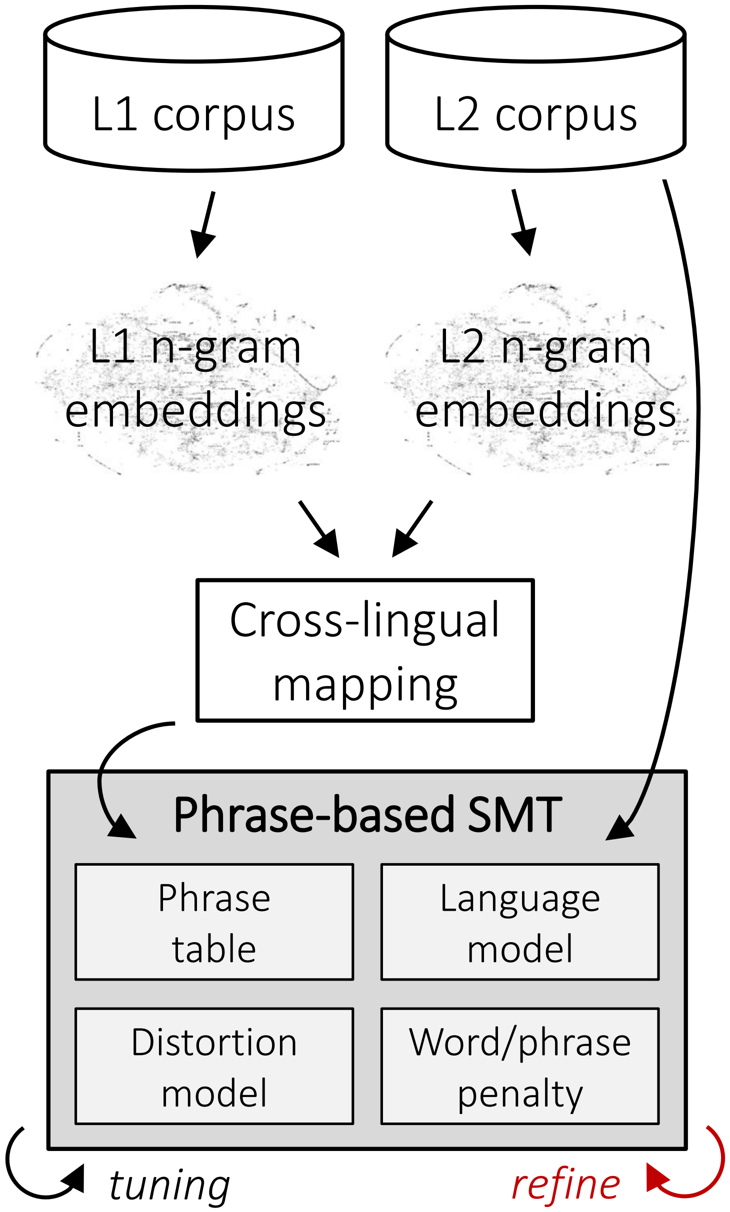
Refinement


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

Refinement

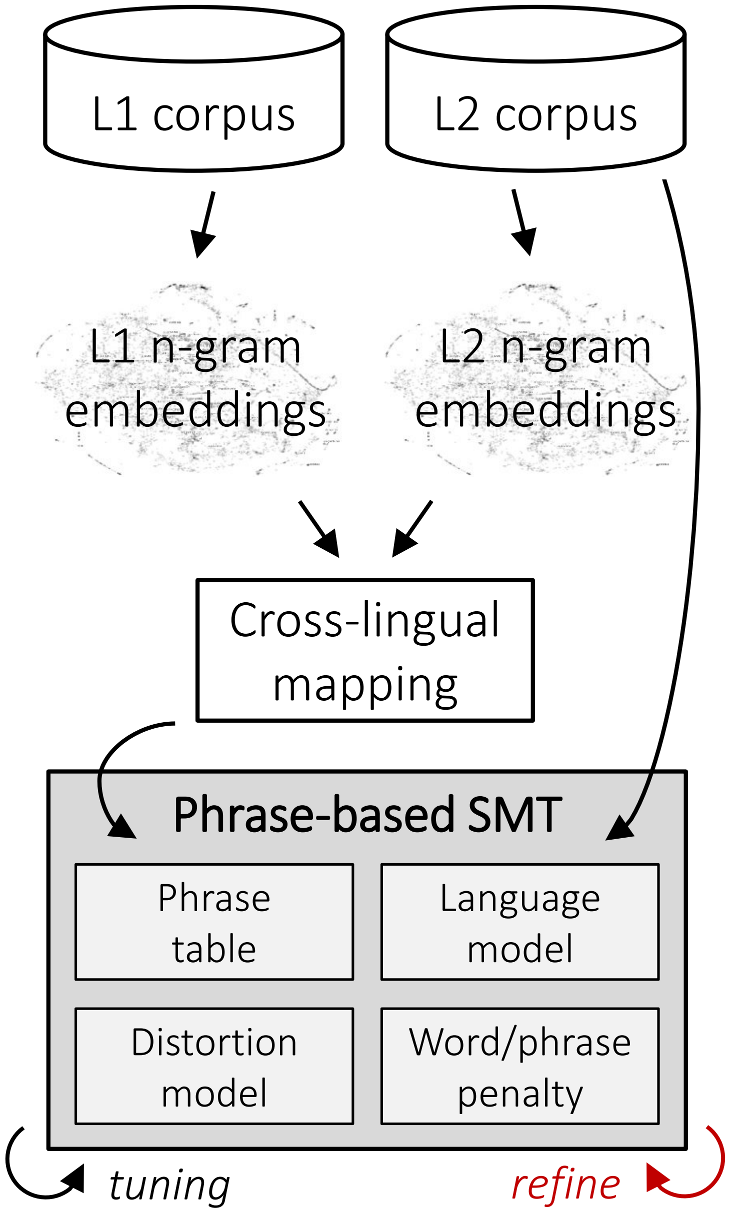

L1 train



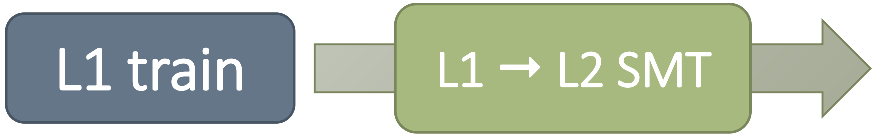
Refinement

L1 train

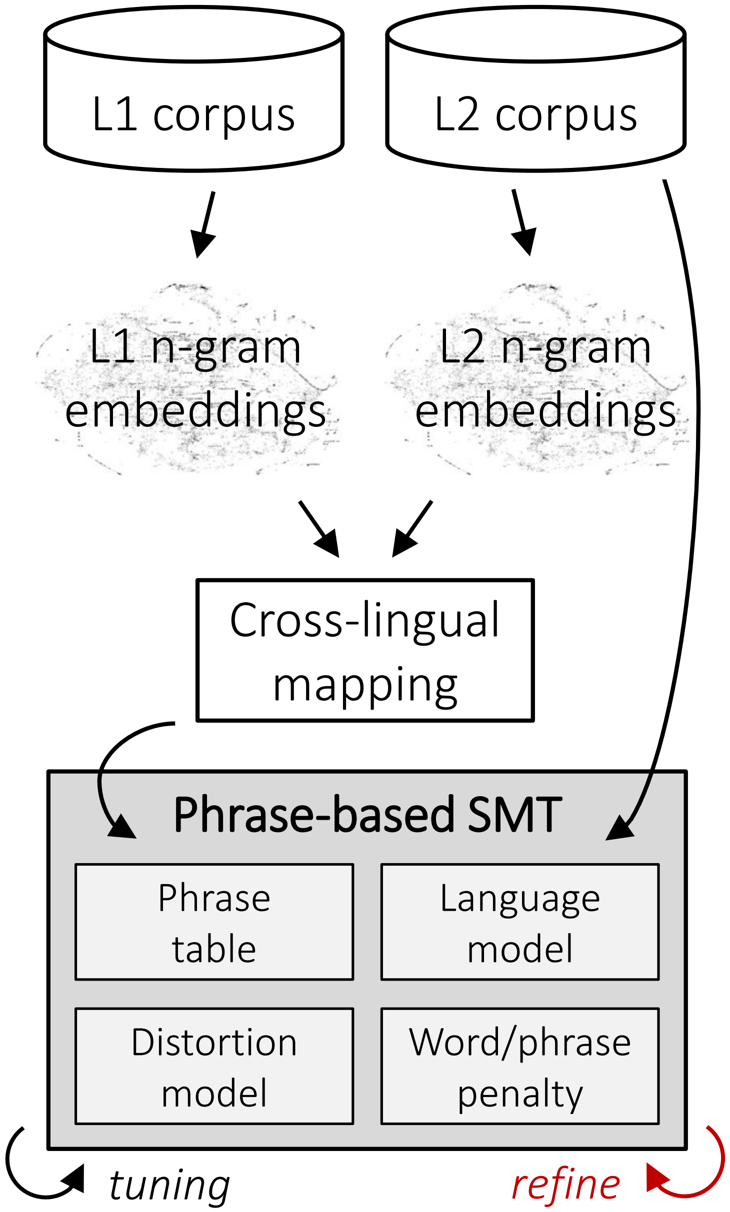
L1 → L2 SMT

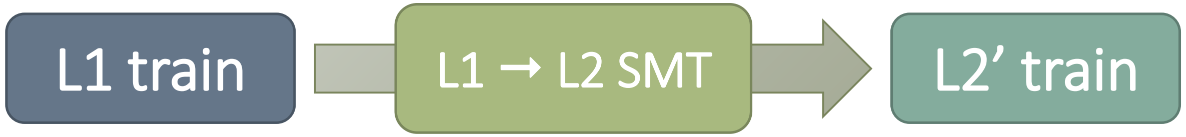

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

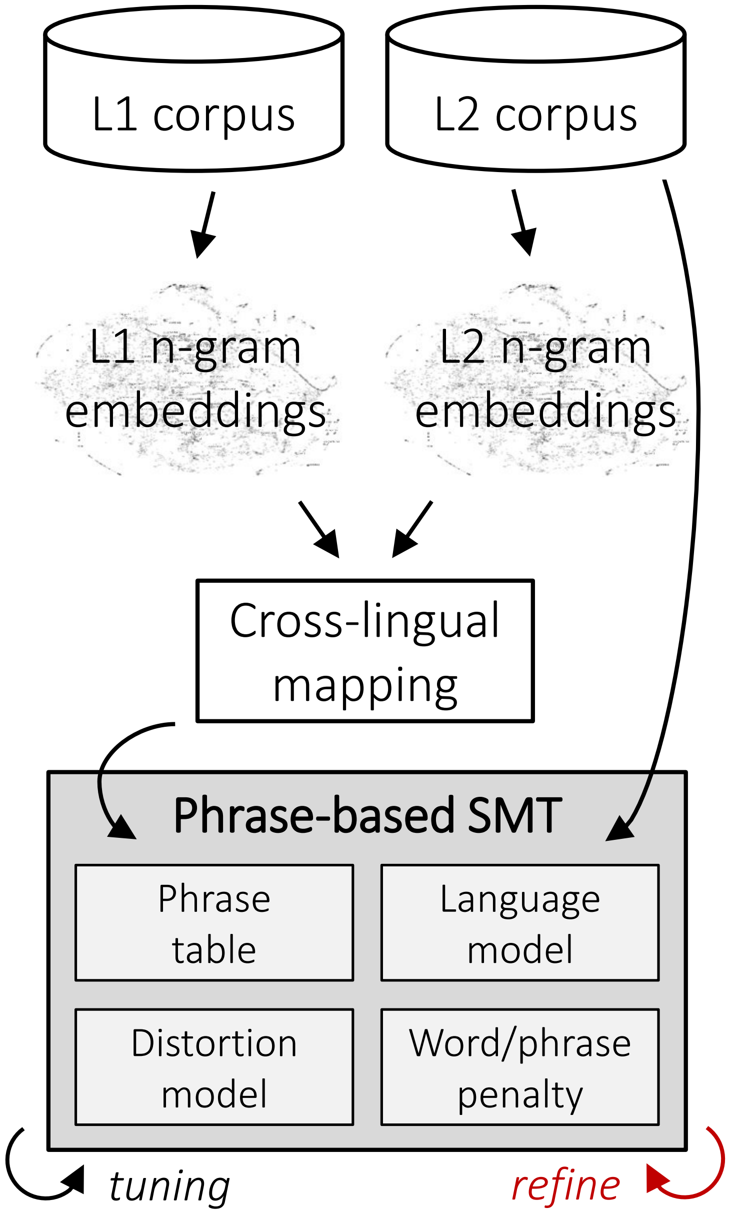
Refinement


L1 train
L1 → L2 SMT

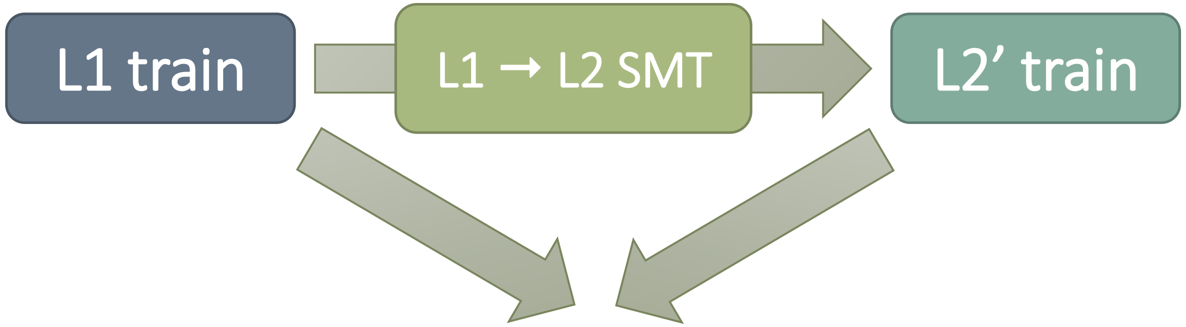
Refinement

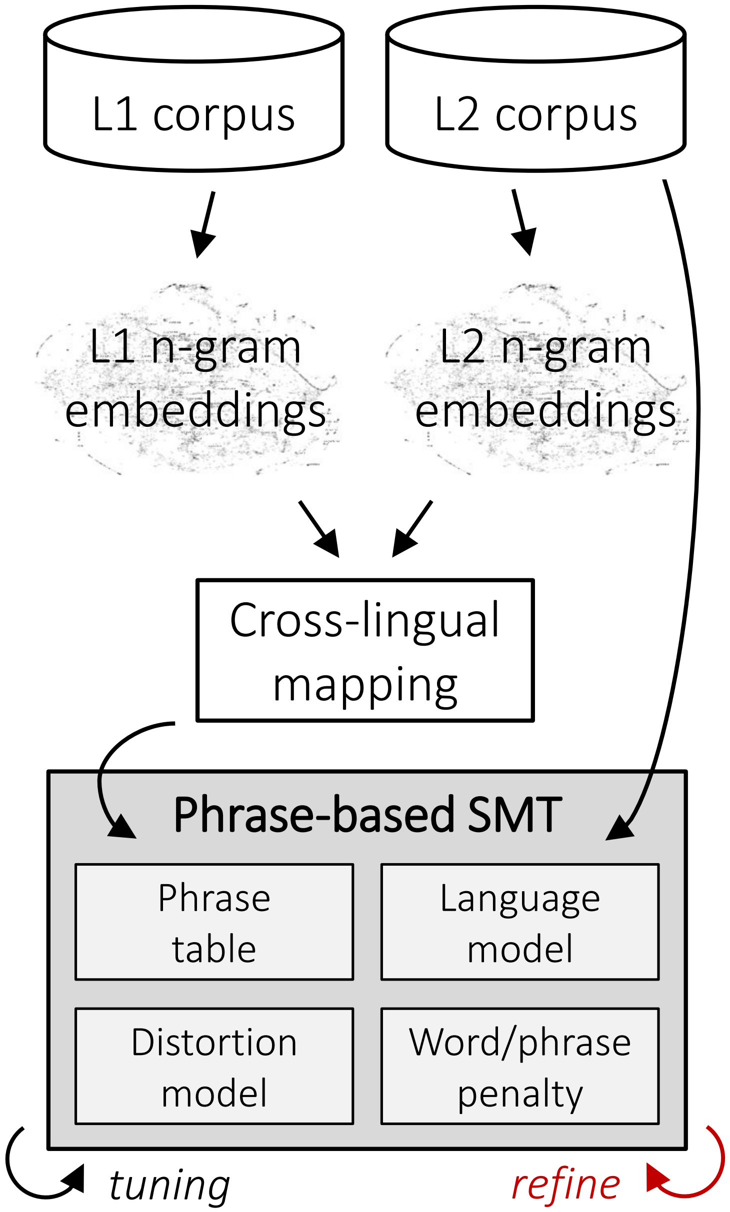

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine


L1 train
L1 → L2 SMT
L2' train

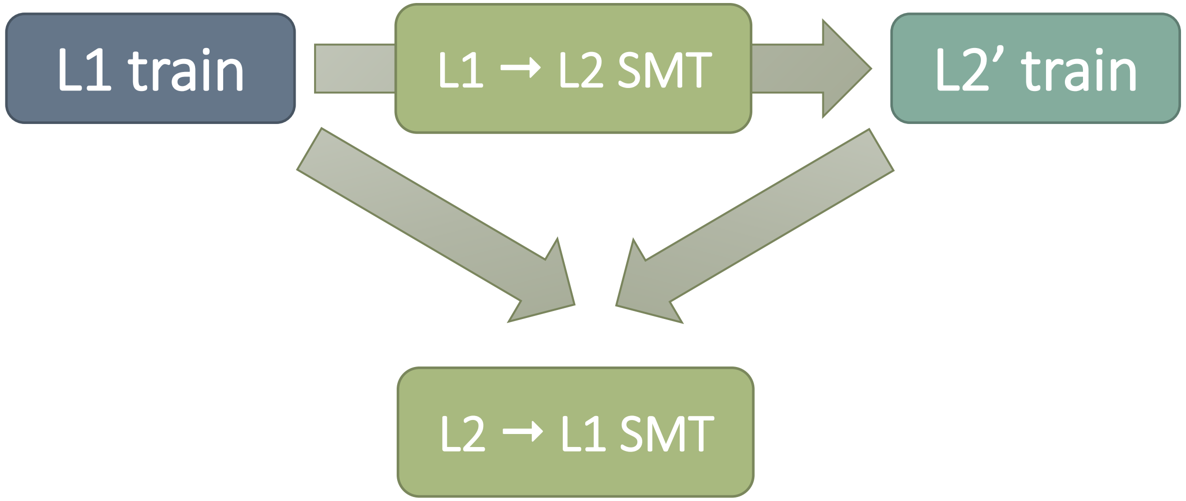

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

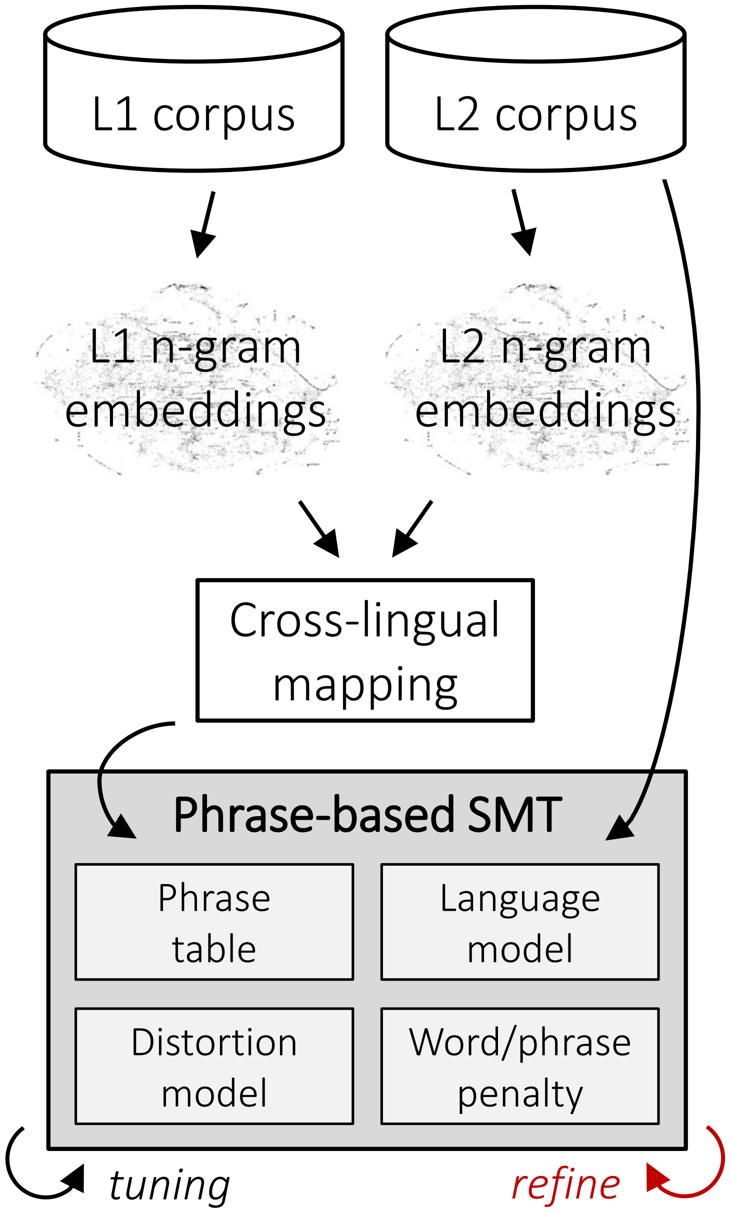
Refinement


L1 train
L1 → L2 SMT
L2' train

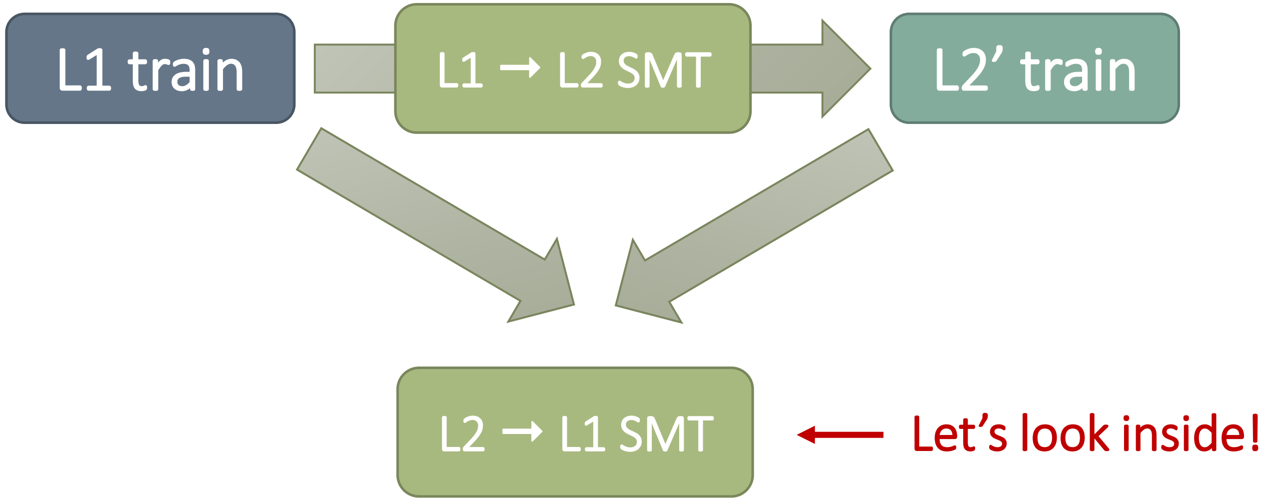

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

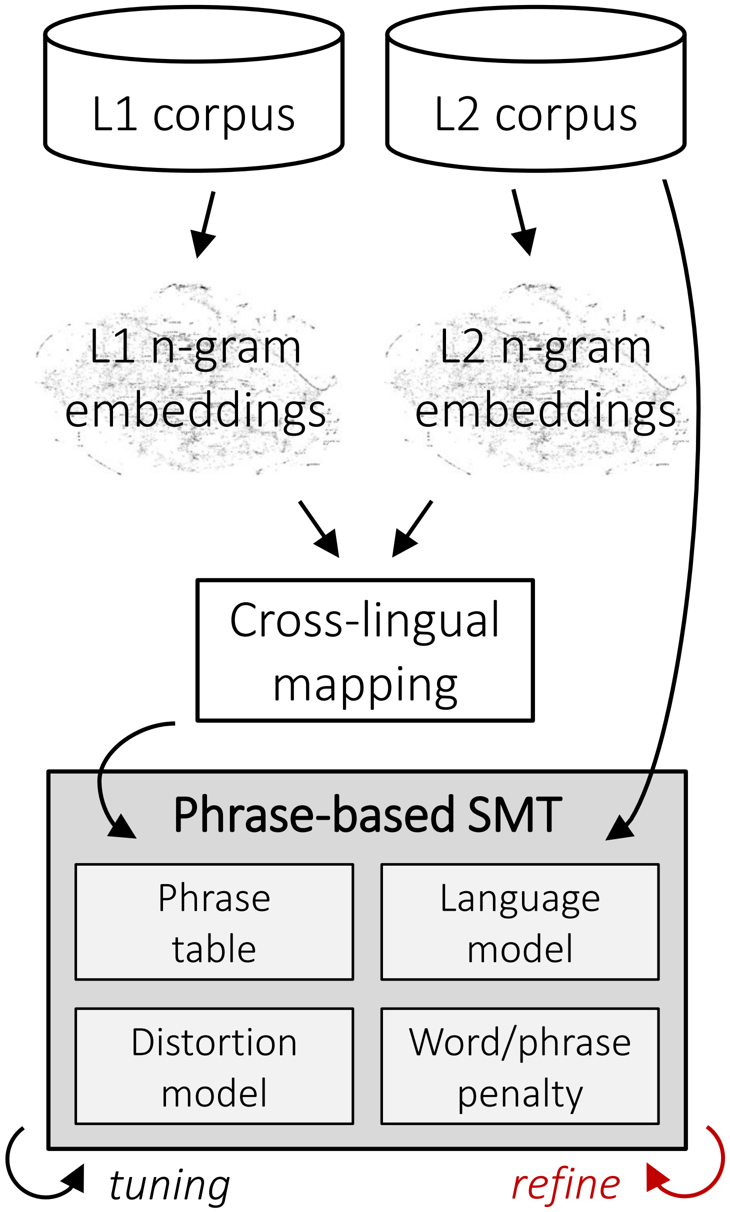
Refinement


L1 train
L1 → L2 SMT
L2’ train
L2 → L1 SMT


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

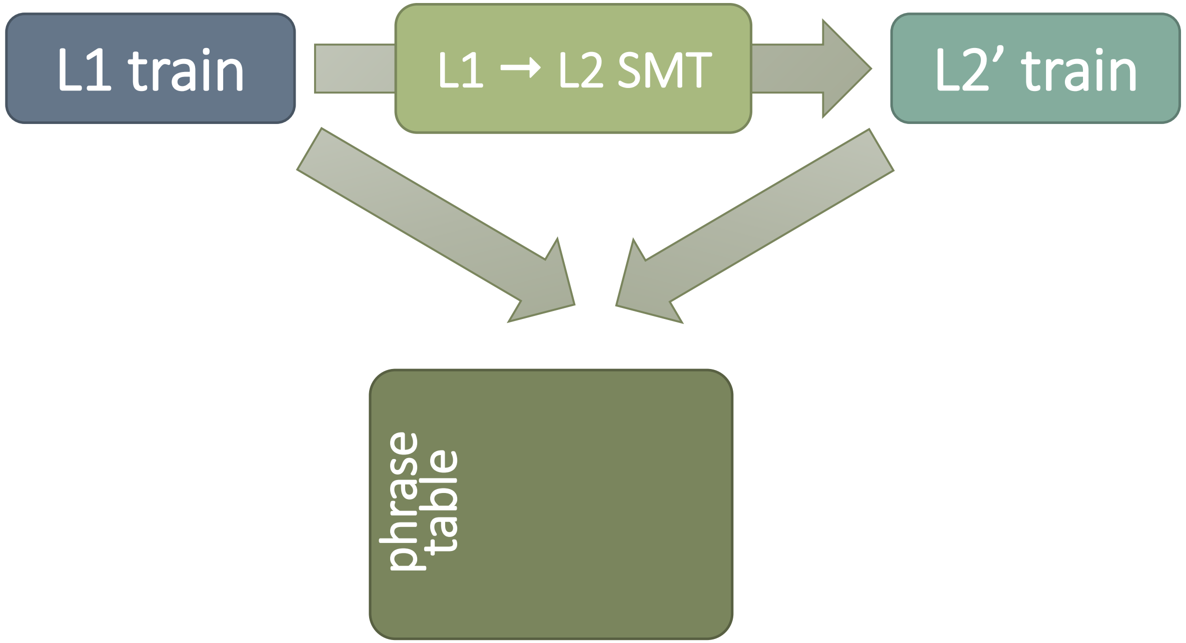
Refinement


L1 train
L1 → L2 SMT
L2' train
L2 → L1 SMT
Let's look inside!

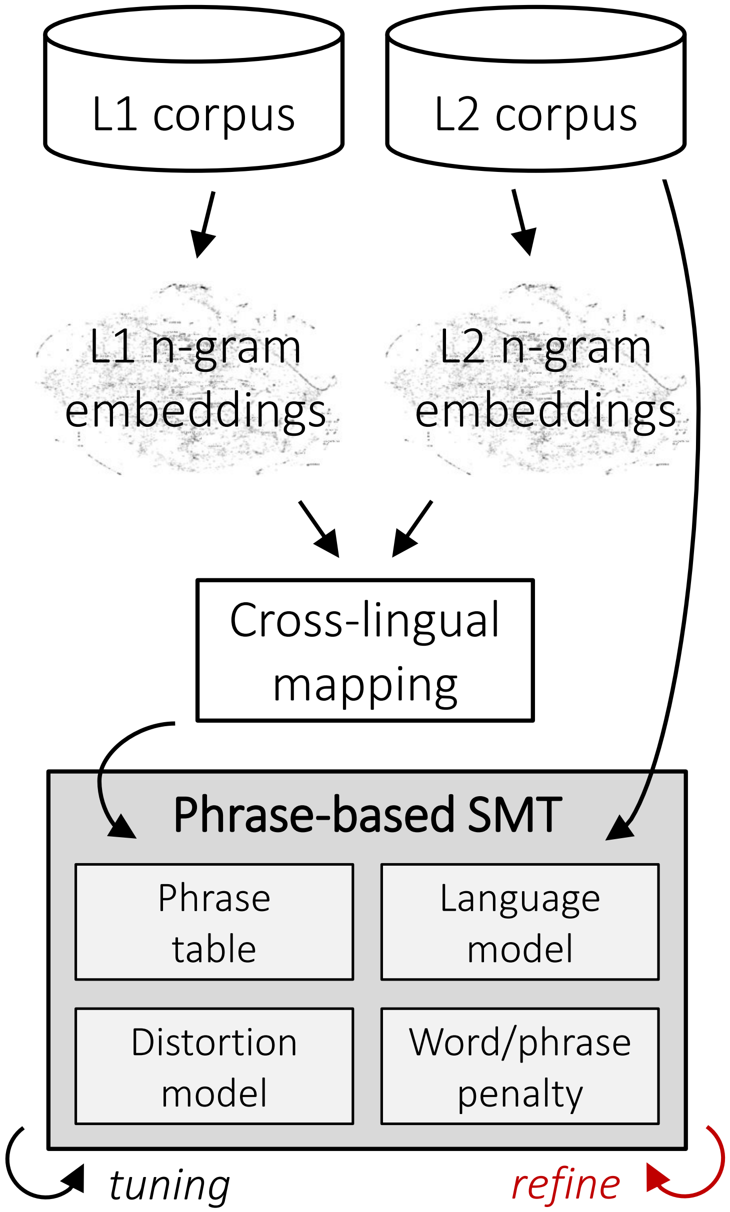


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

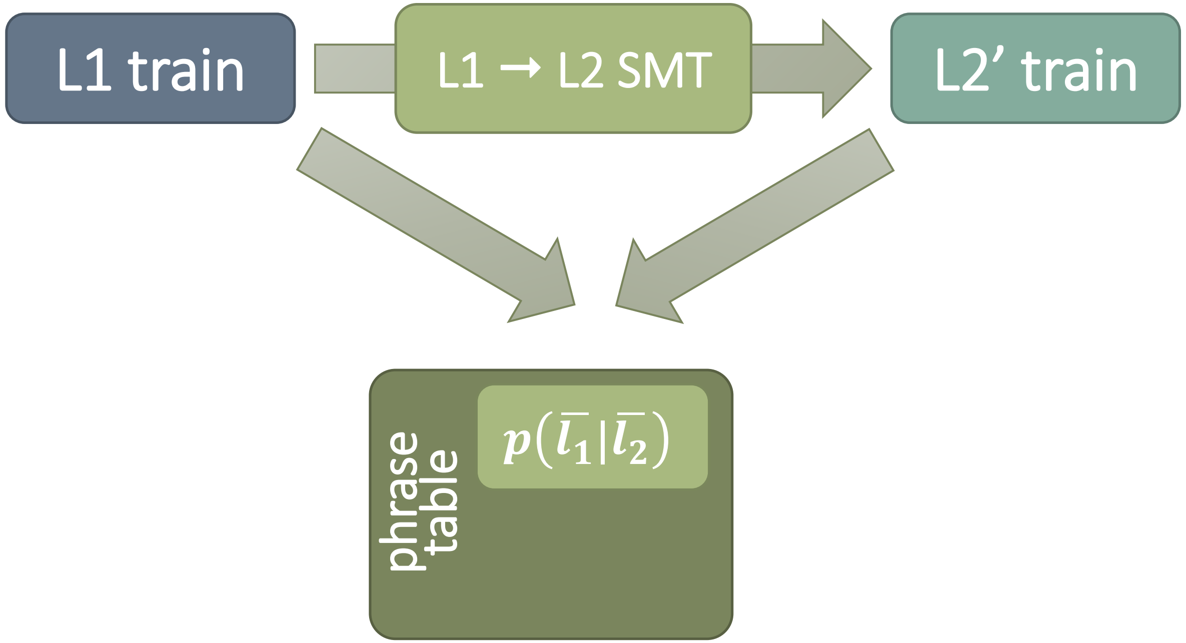
Refinement

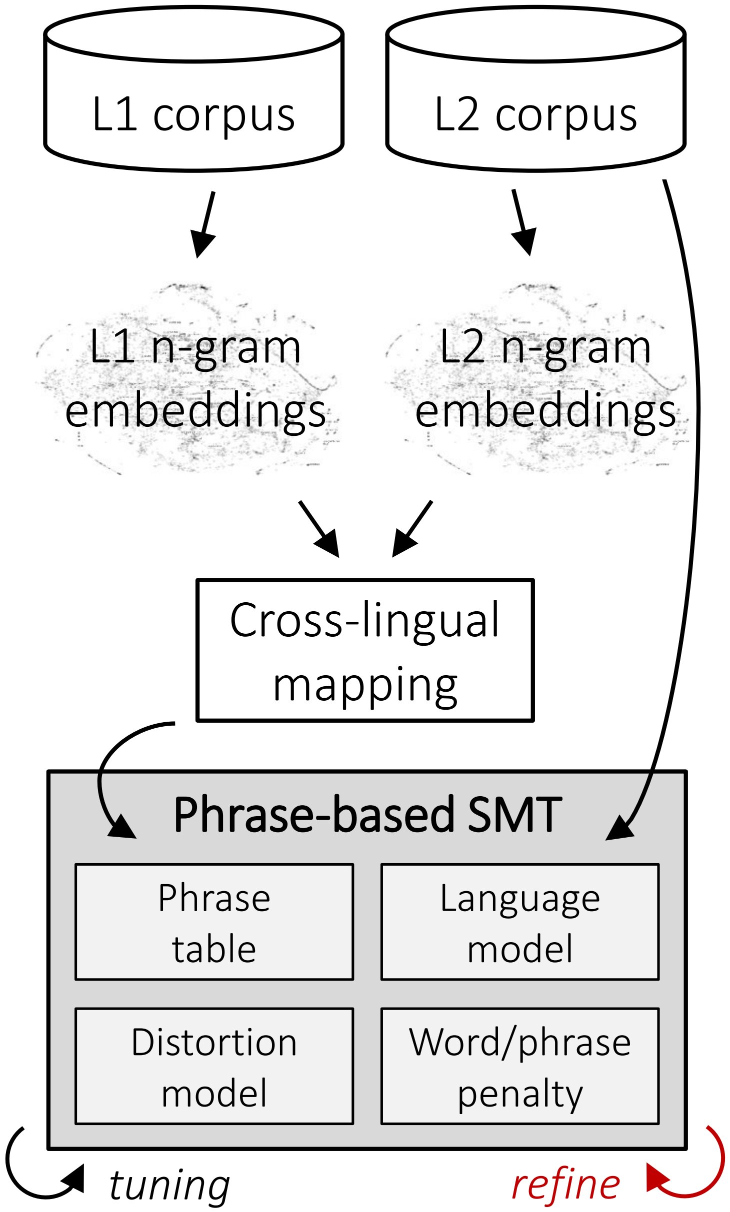


L1 train
L1 → L2 SMT
L2’ train
phrase table

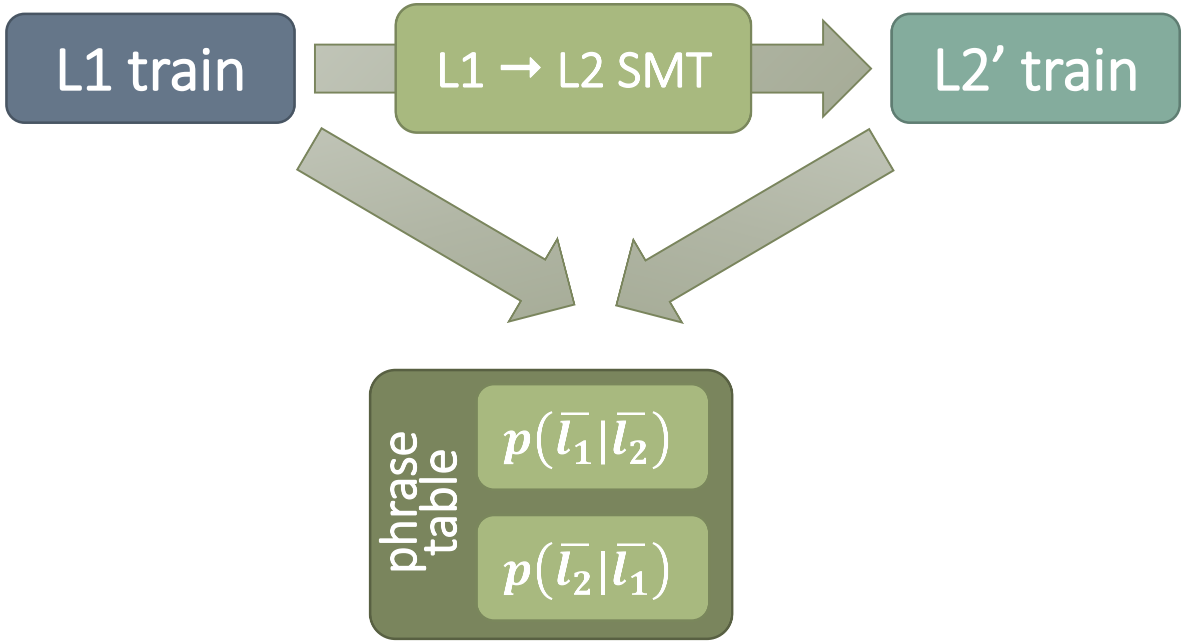

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

Refinement

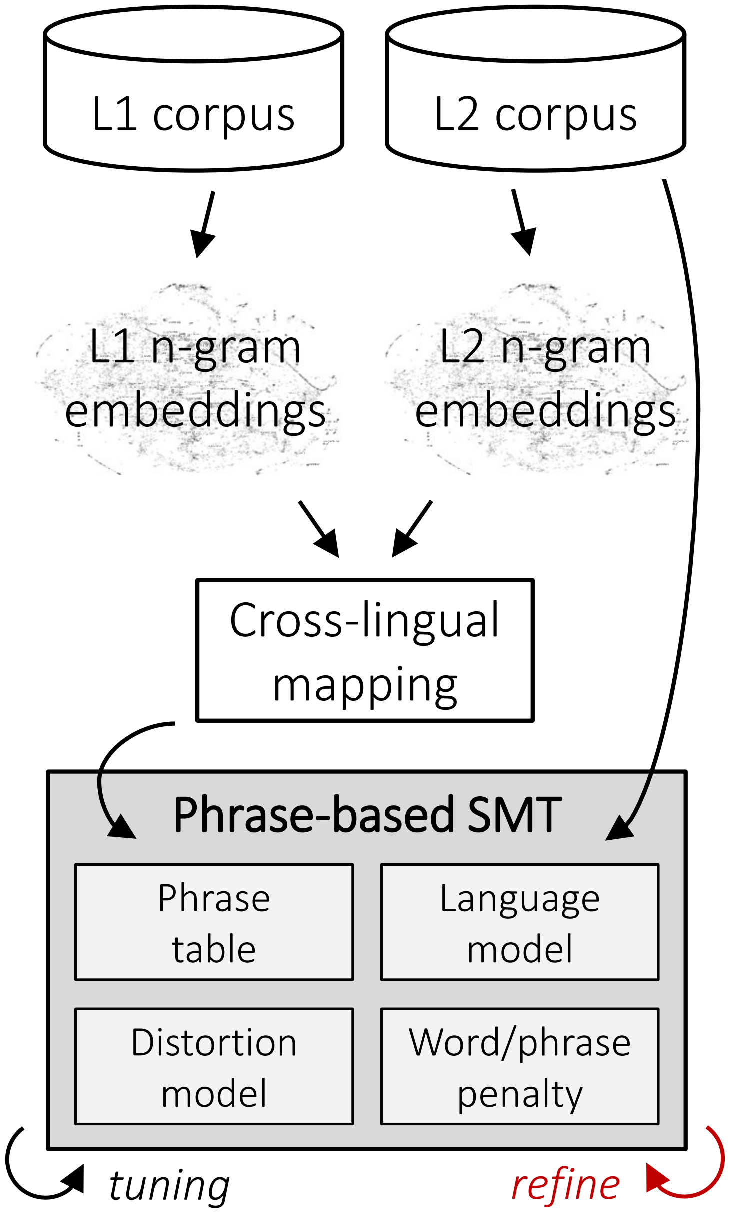

L1 train
L1 → L2 SMT
L2' train
phrase table
$p(\bar{l_1}|\bar{l_2})$

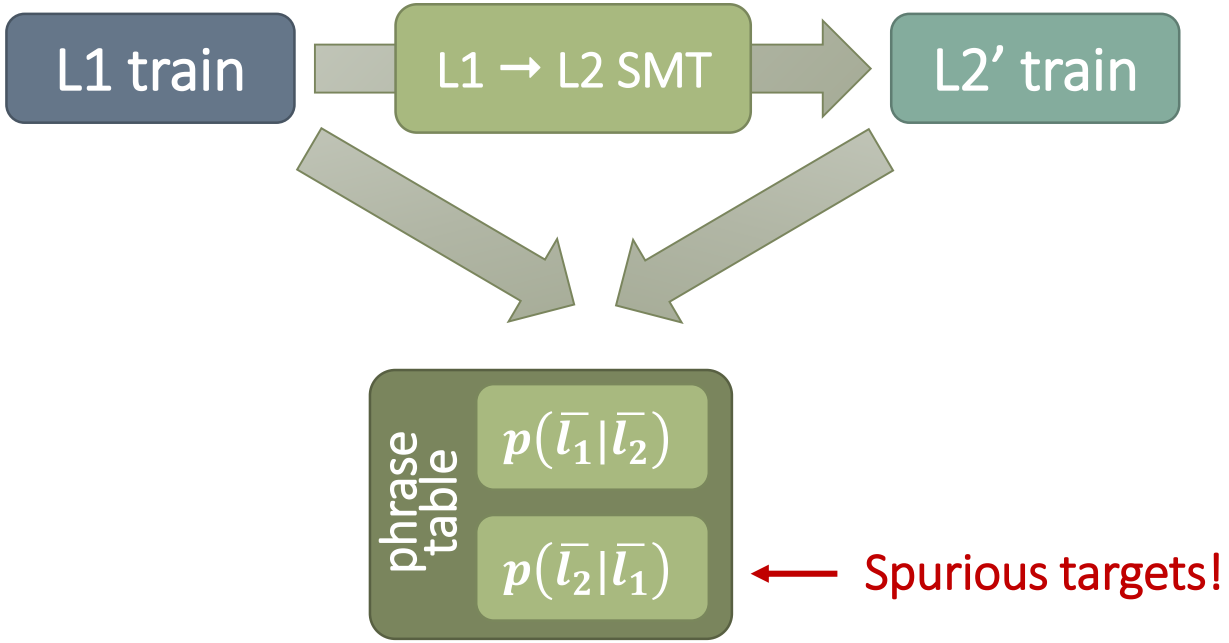

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

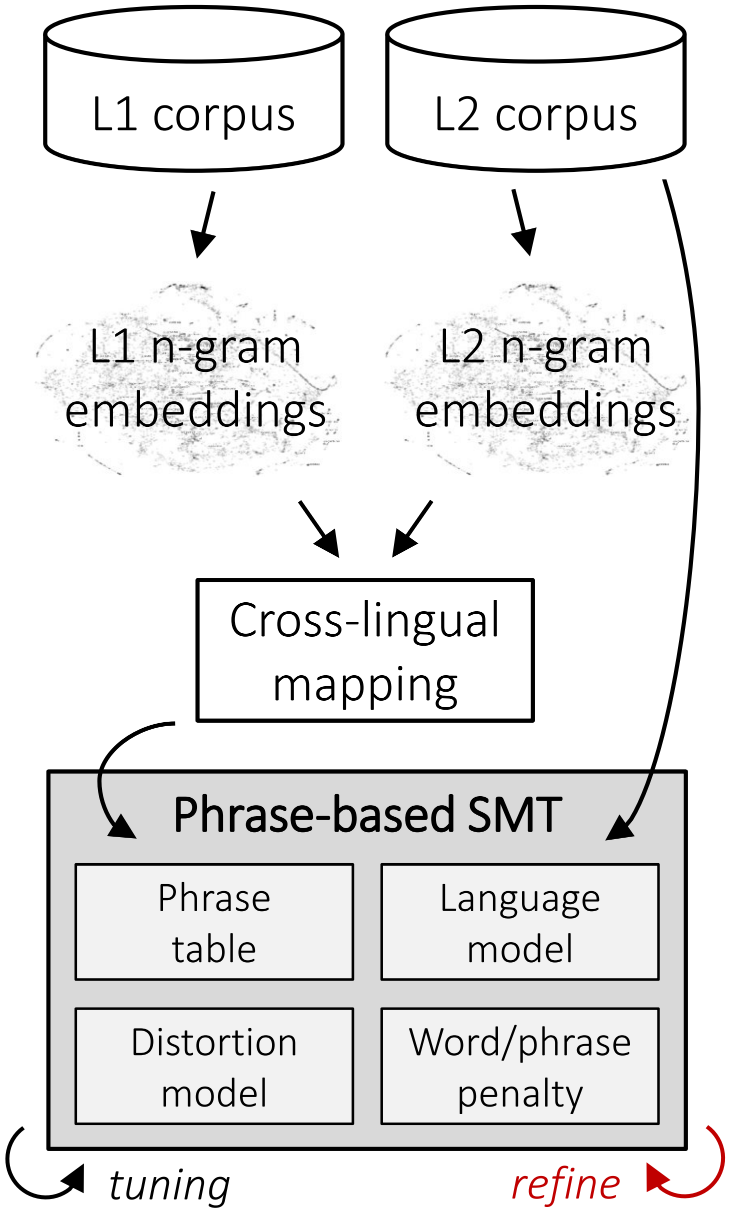
Refinement


L1 train
L1 → L2 SMT
L2’ train
phrase table
$p(\bar{l_1}|\bar{l_2})$
$p(\bar{l_2}|\bar{l_1})$

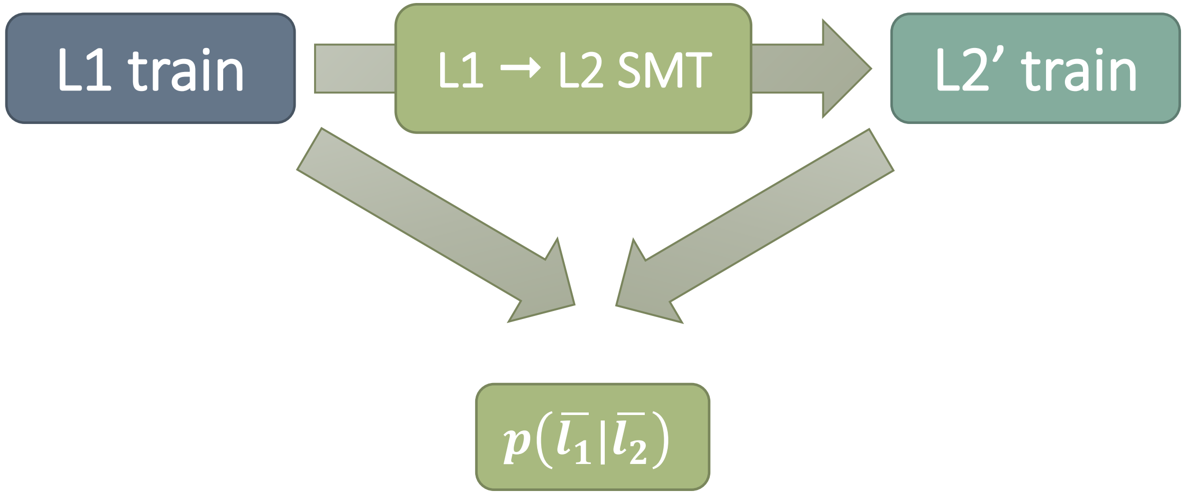
Refinement


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

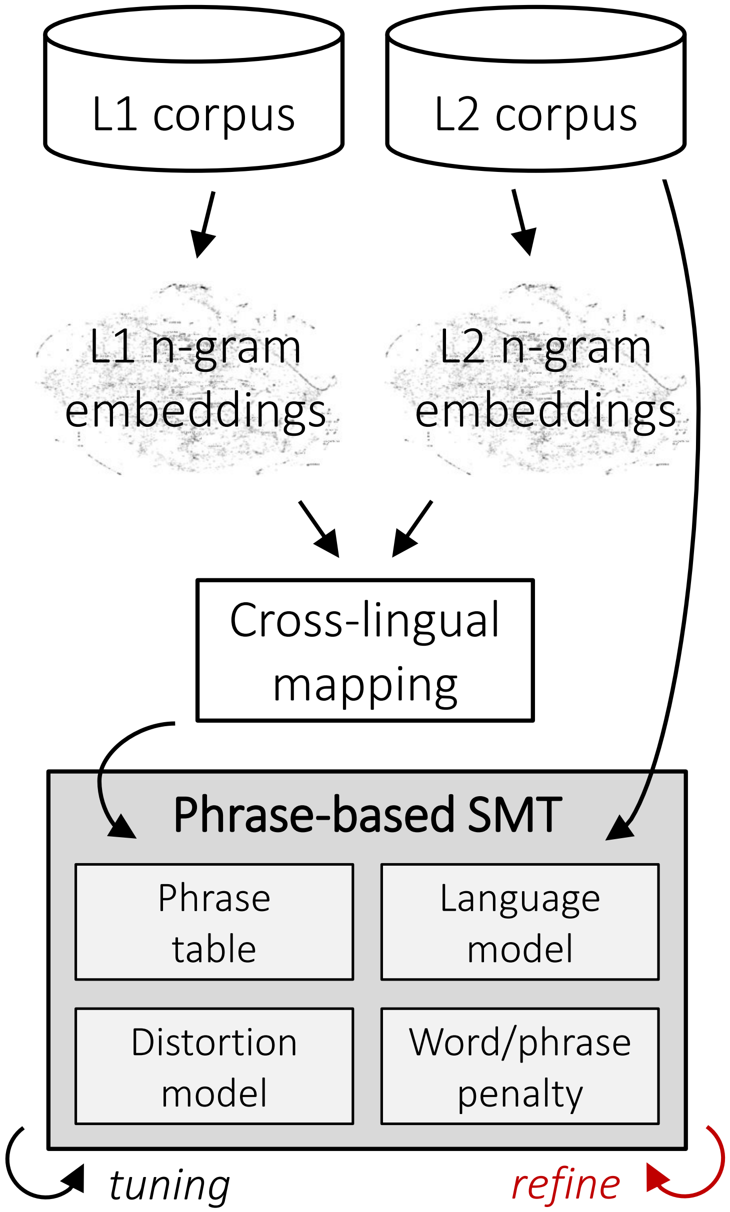

L1 train
L1 → L2 SMT
L2' train
phrase table
$p(\bar{l_1}|\bar{l_2})$
$p(\bar{l_2}|\bar{l_1})$
Spurious targets!

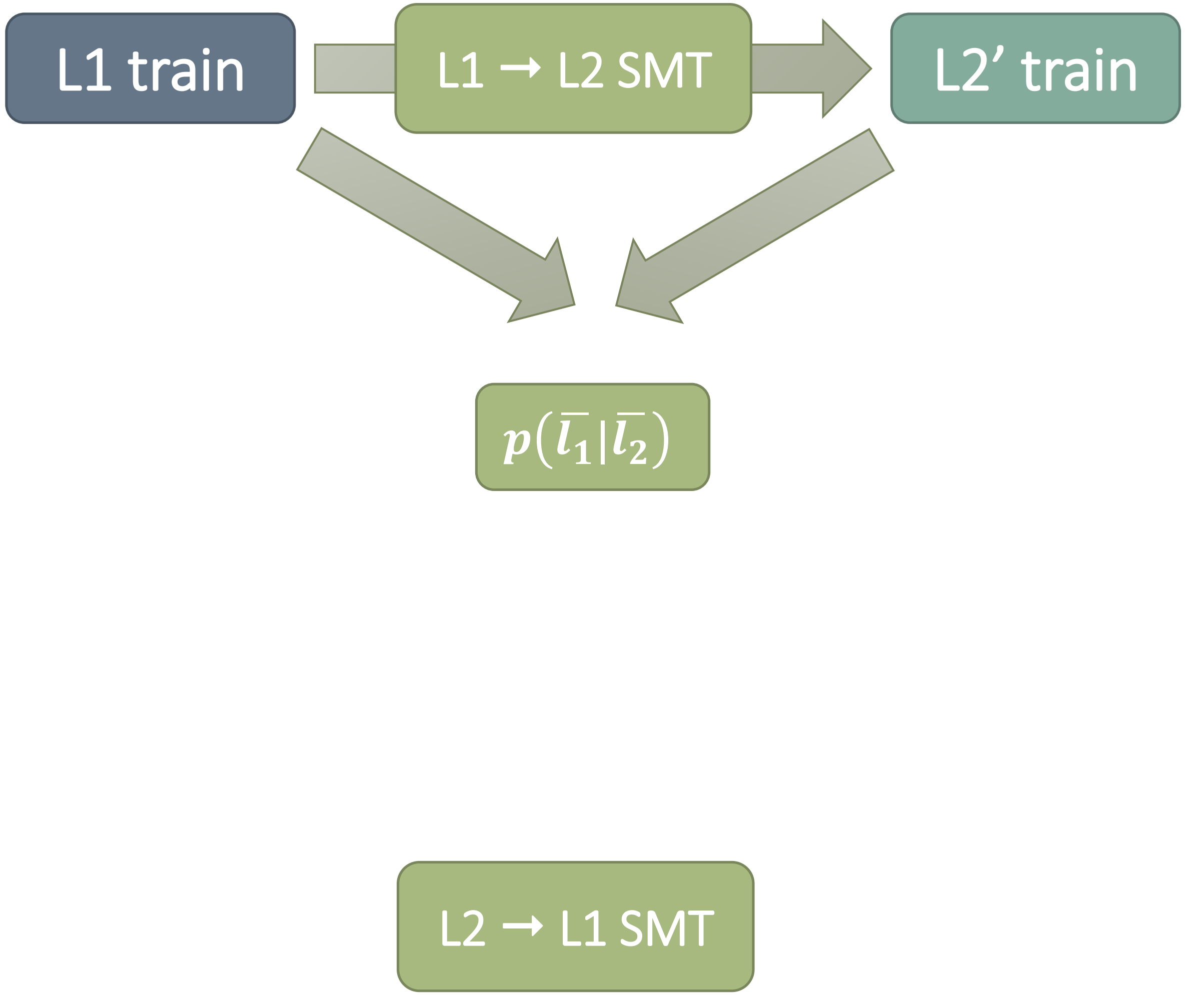

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

Refinement

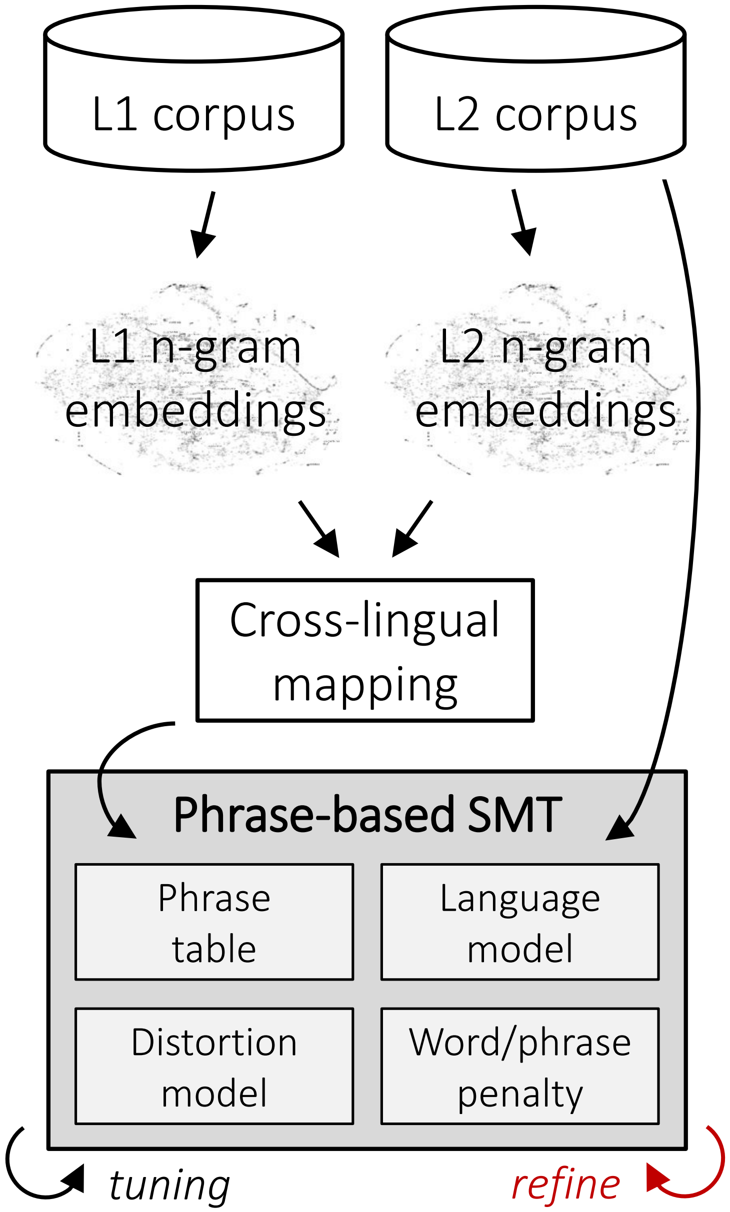

L1 train
L1 → L2 SMT
L2' train
$p(\bar{l_1}|\bar{l_2})$

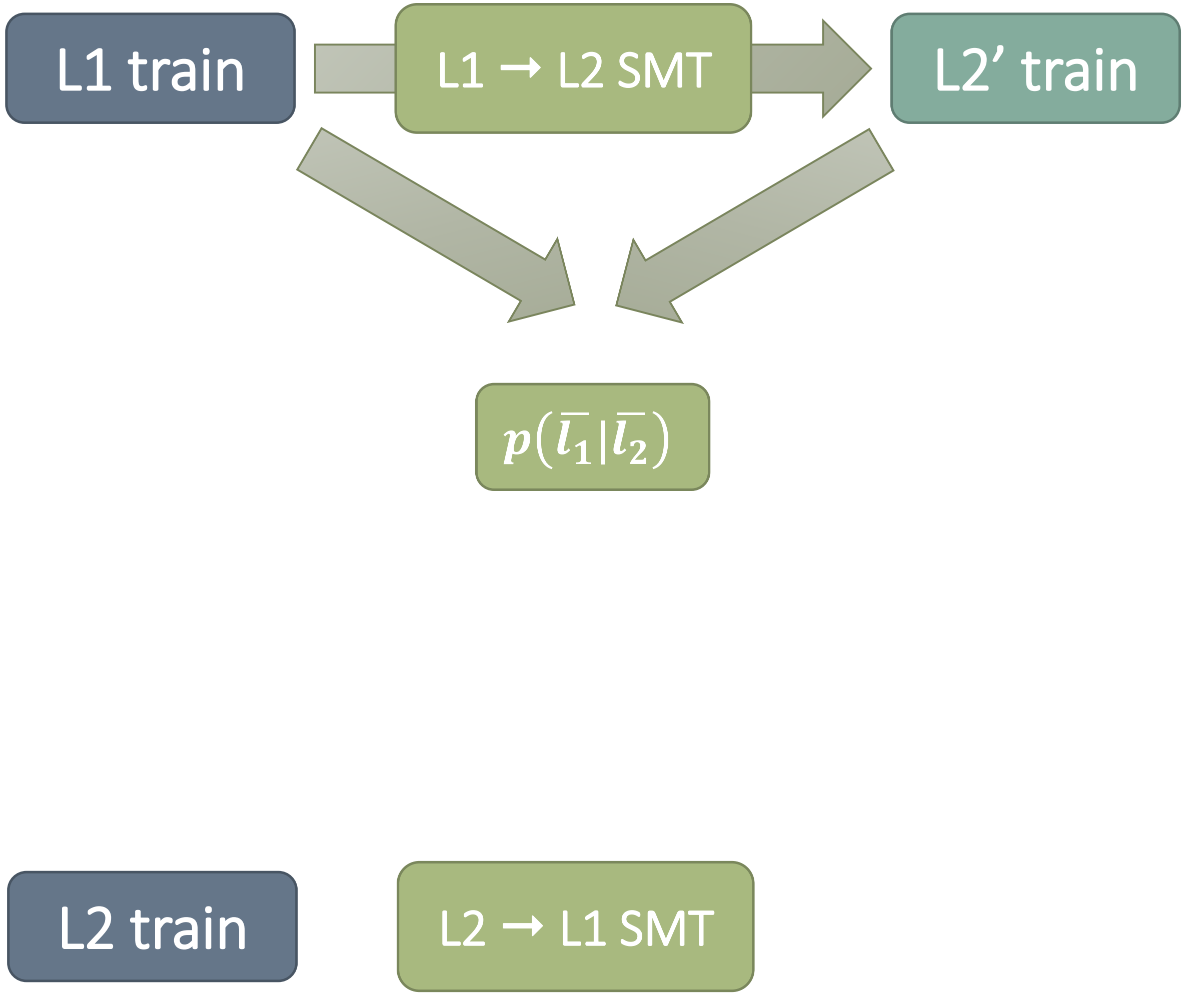
Refinement


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

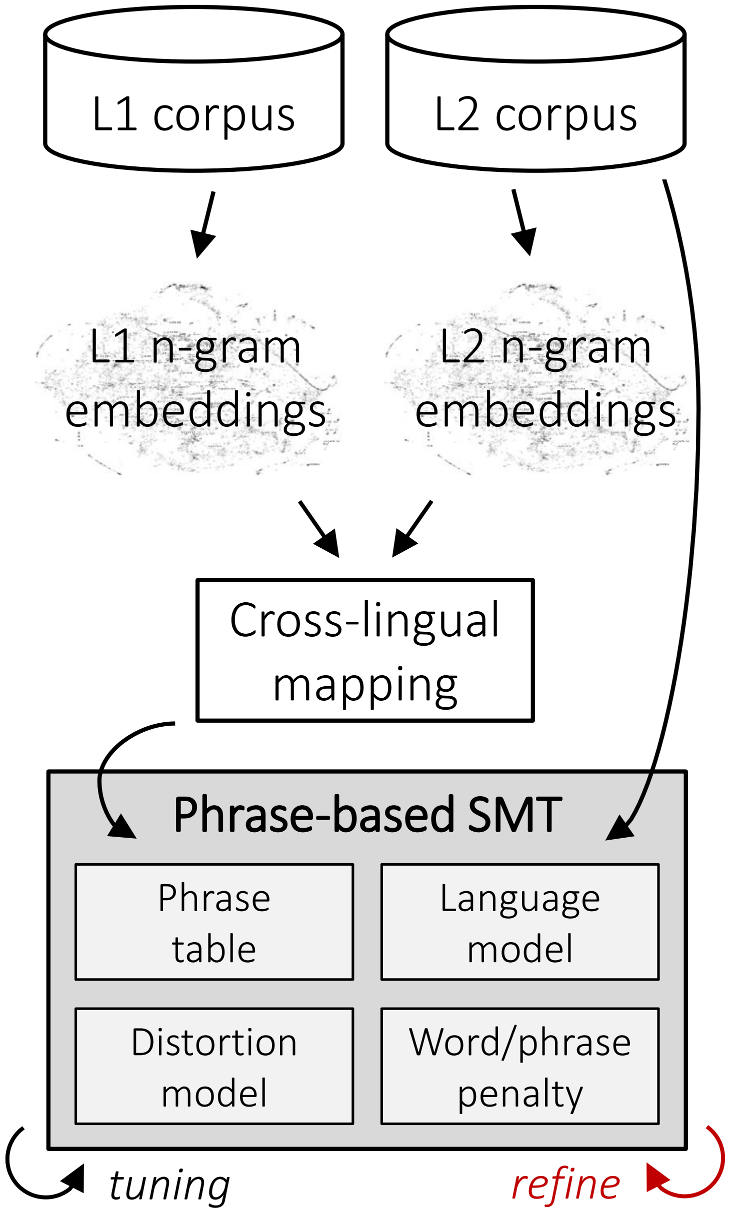

L1 train
L1 → L2 SMT
L2’ train
$p(\bar{l_1}|\bar{l_2})$
L2 → L1 SMT

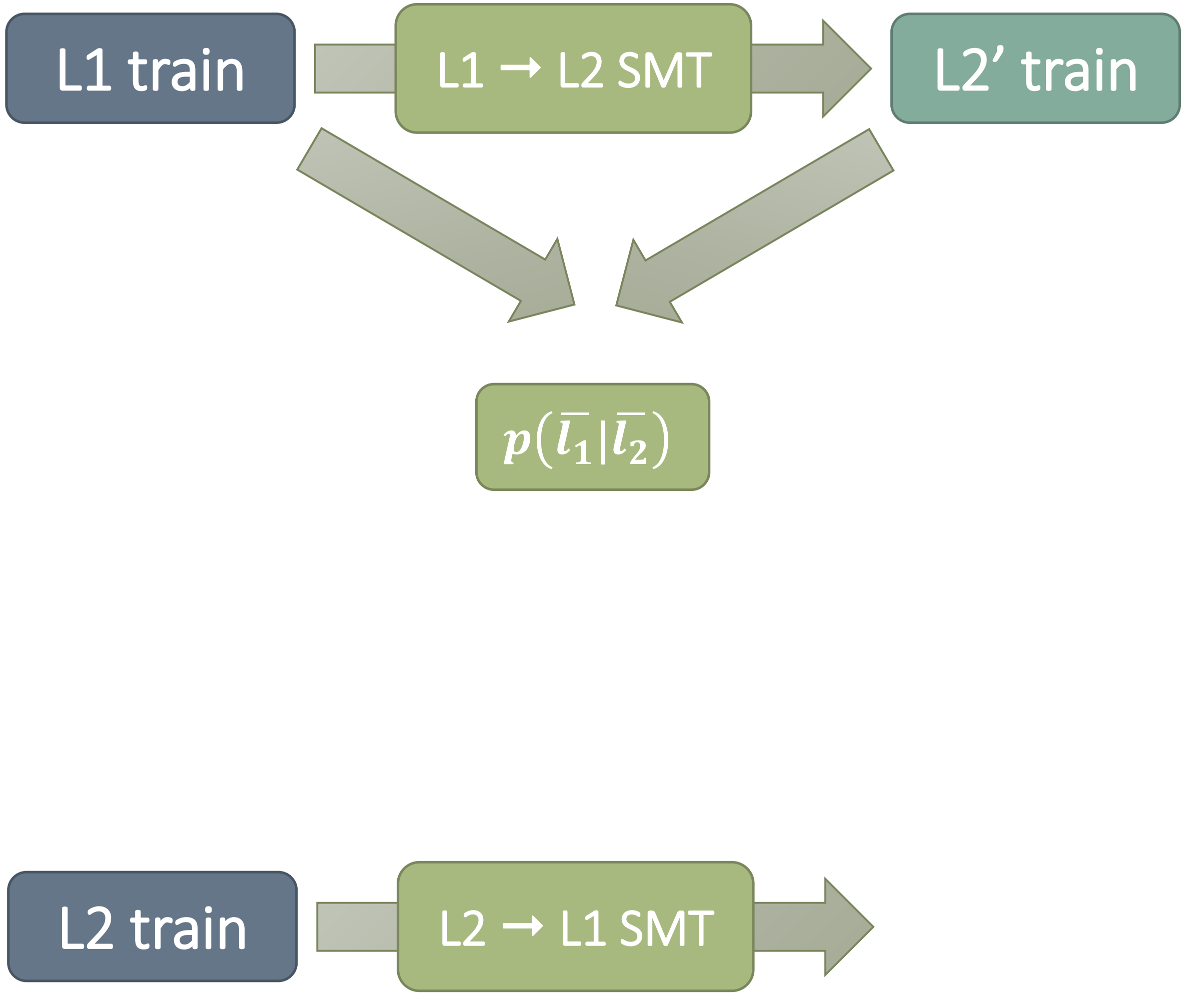
Refinement

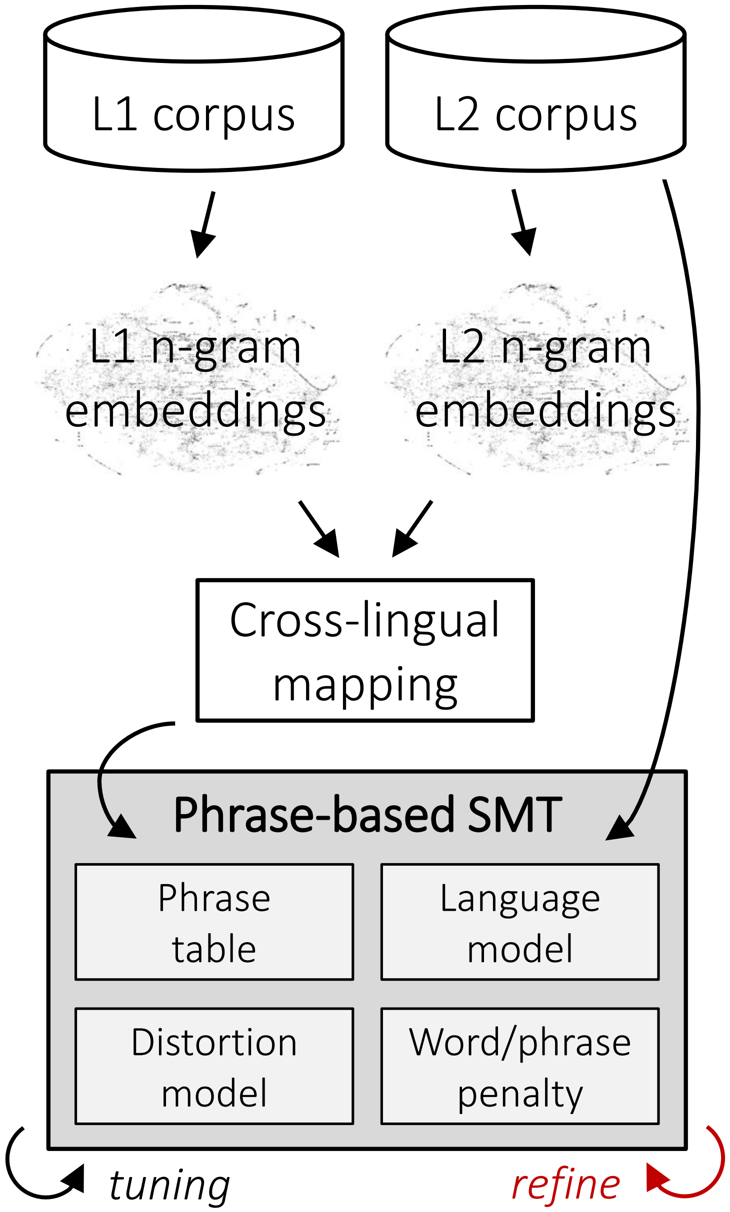

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine


L1 train
L1 → L2 SMT
L2' train
$p(\bar{l_1}|\bar{l_2})$
L2 train
L2 → L1 SMT

Refinement

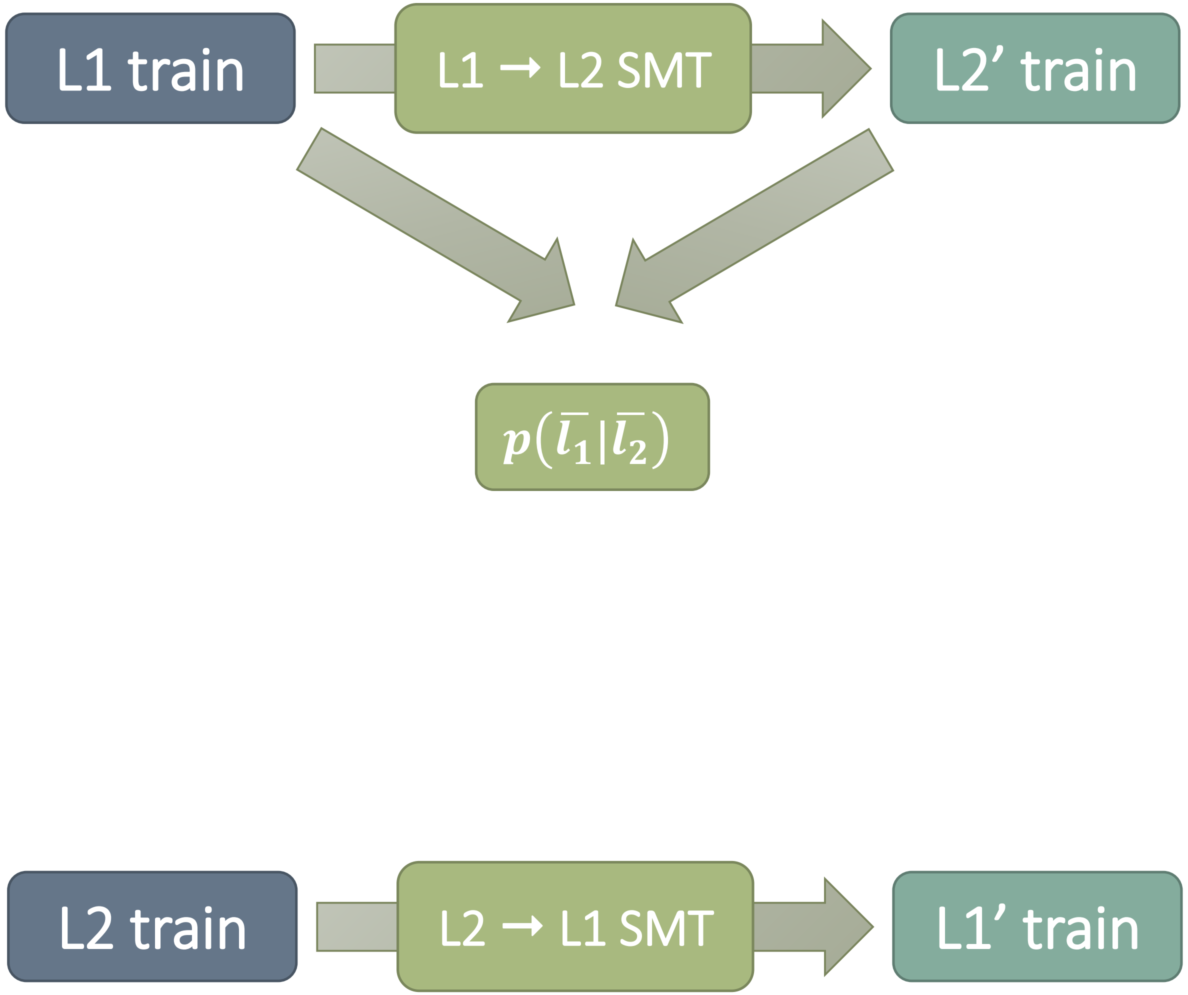

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine


L1 train
L1 → L2 SMT
L2' train
$p(\bar{l_1}|\bar{l_2})$
L2 train
L2 → L1 SMT

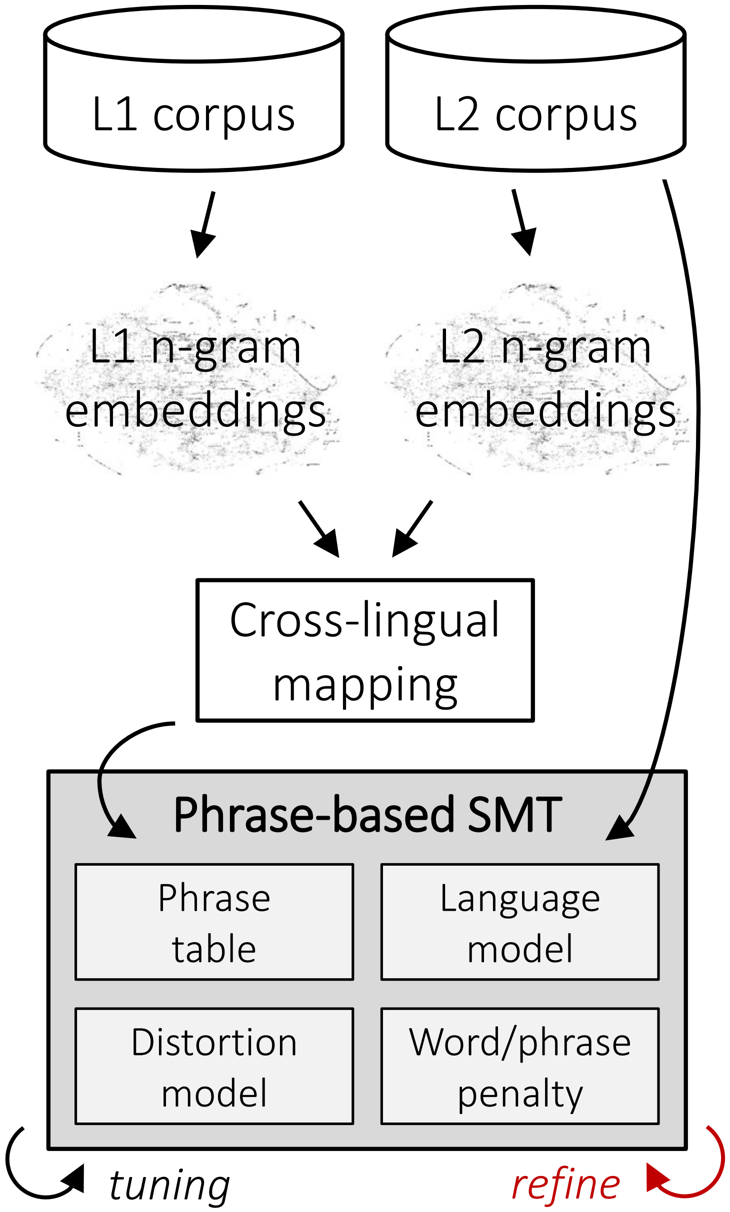


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

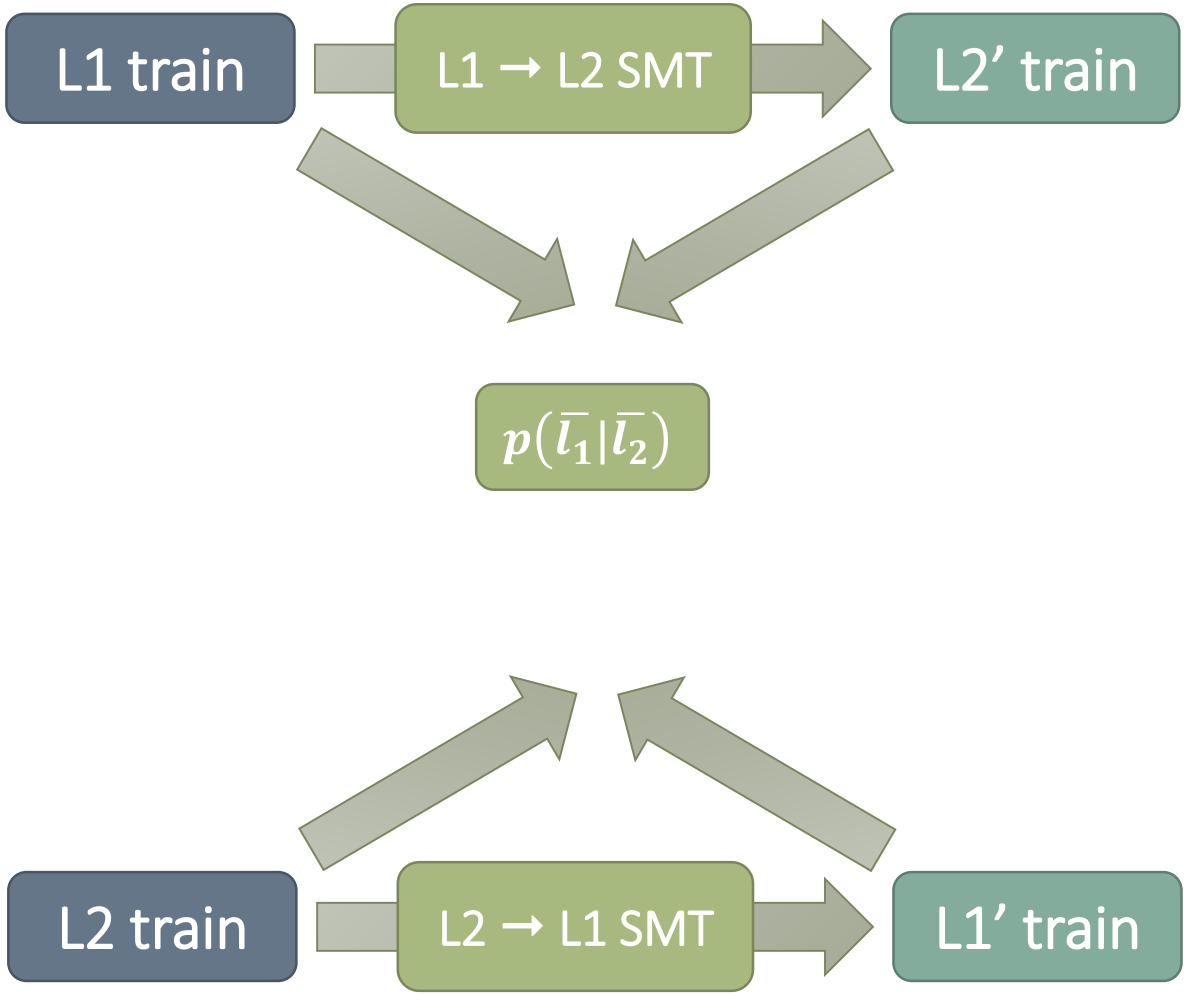
Refinement

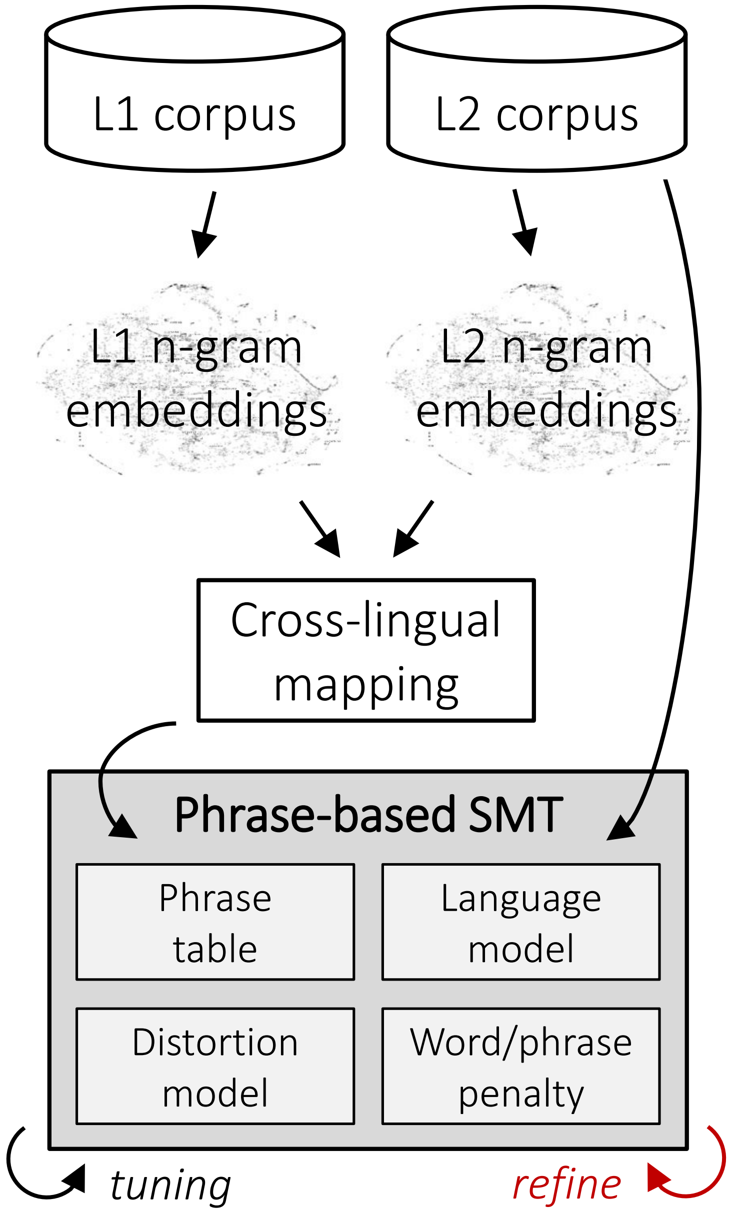


L1 train
L1 → L2 SMT
L2' train
$p(\bar{l_1}|\bar{l_2})$
L2 train
L2 → L1 SMT
L1' train

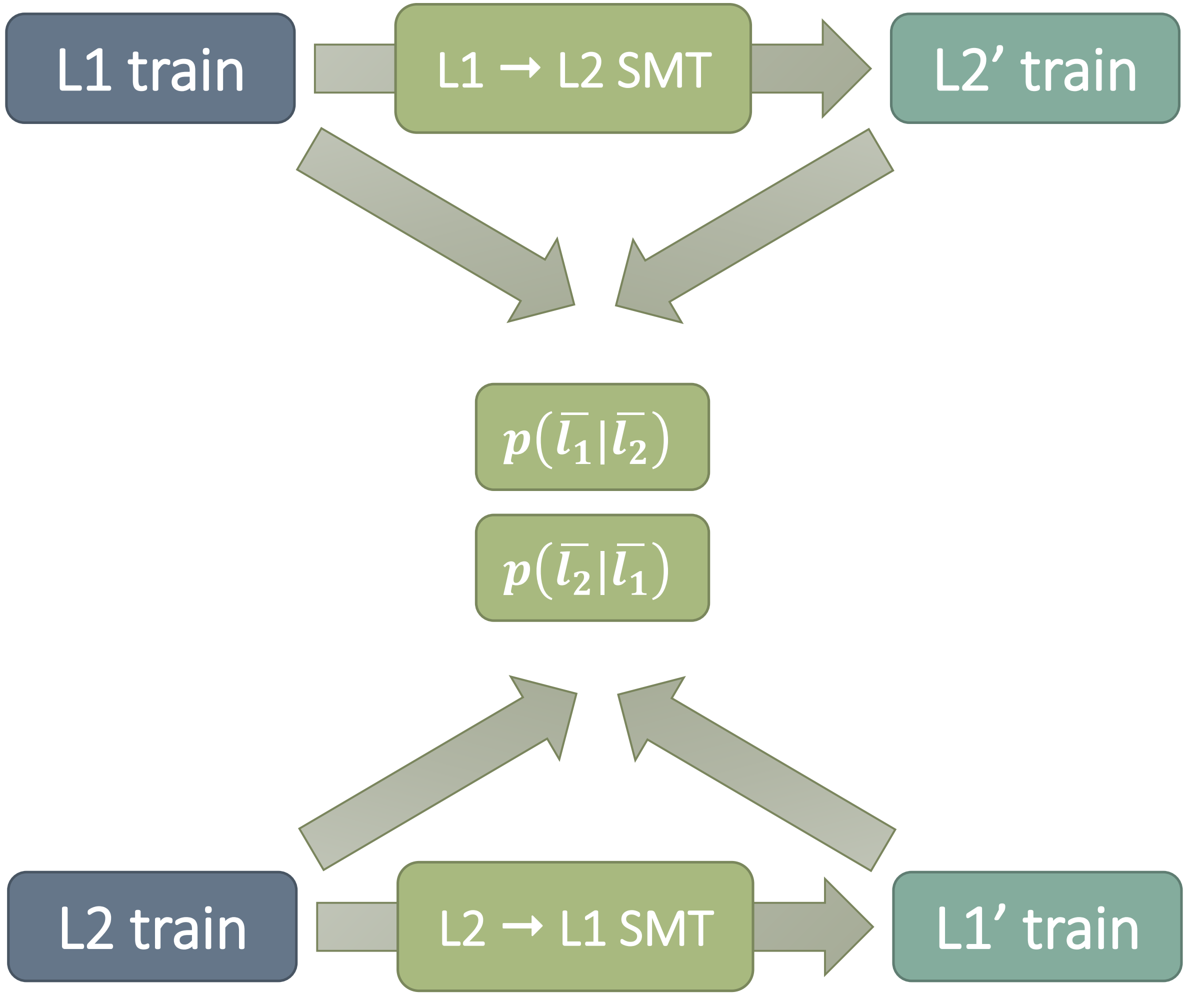

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

Refinement

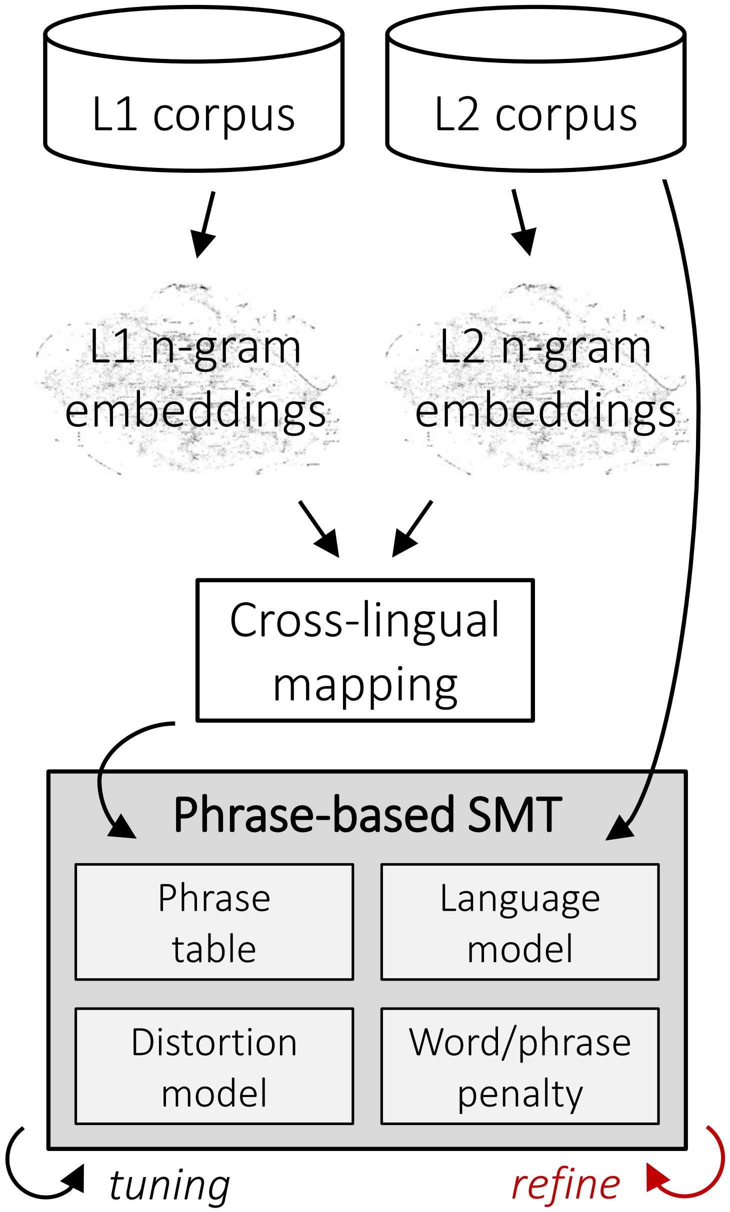

L1 train
L1 → L2 SMT
L2' train
$p(\bar{l_1}|\bar{l_2})$
L2 train
L2 → L1 SMT
L1' train


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

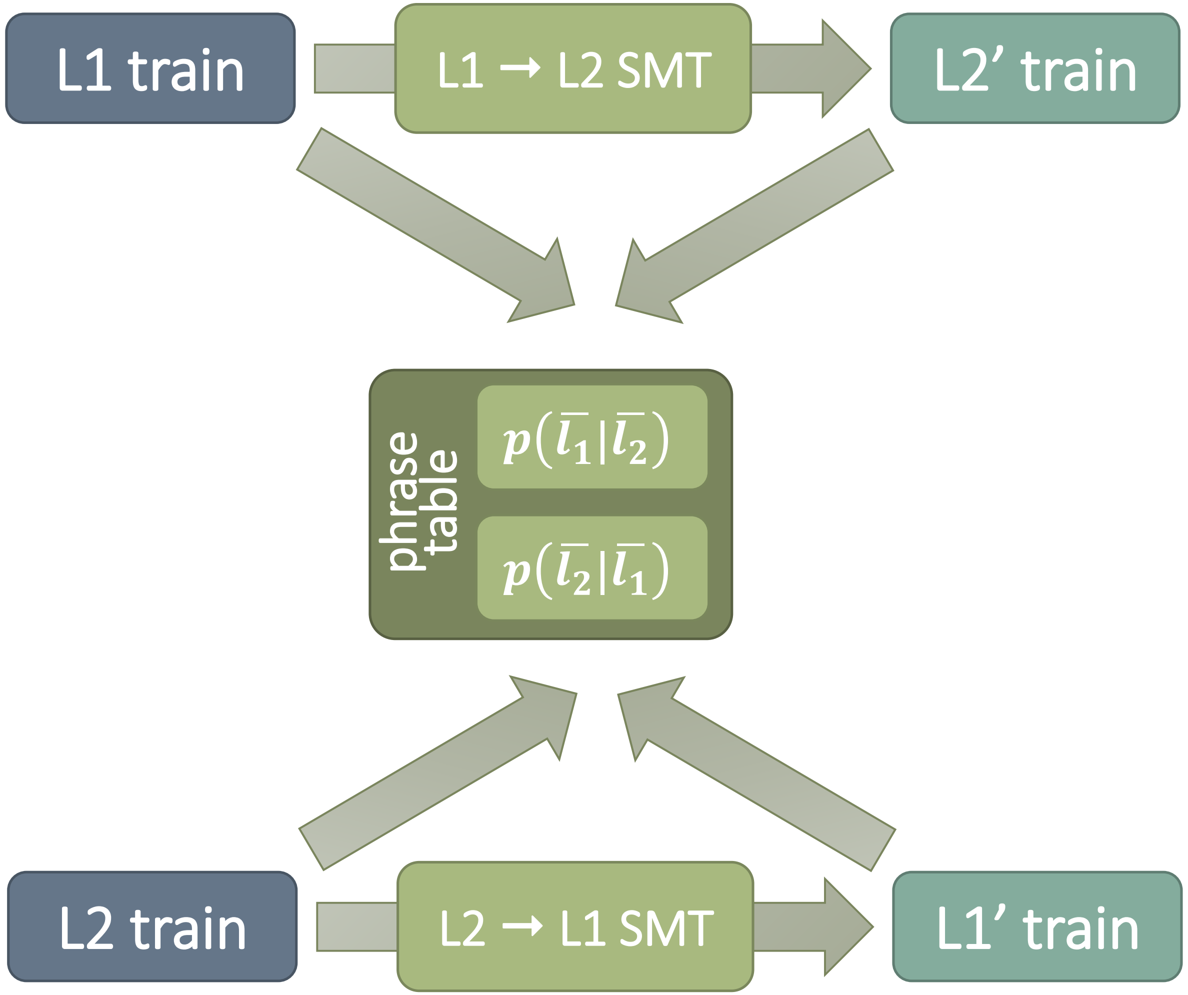
Refinement


L1 train
L1 → L2 SMT
L2' train
$p(\bar{l_1}|\bar{l_2})$
$p(\bar{l_2}|\bar{l_1})$
L2 train
L2 → L1 SMT
L1' train

Refinement

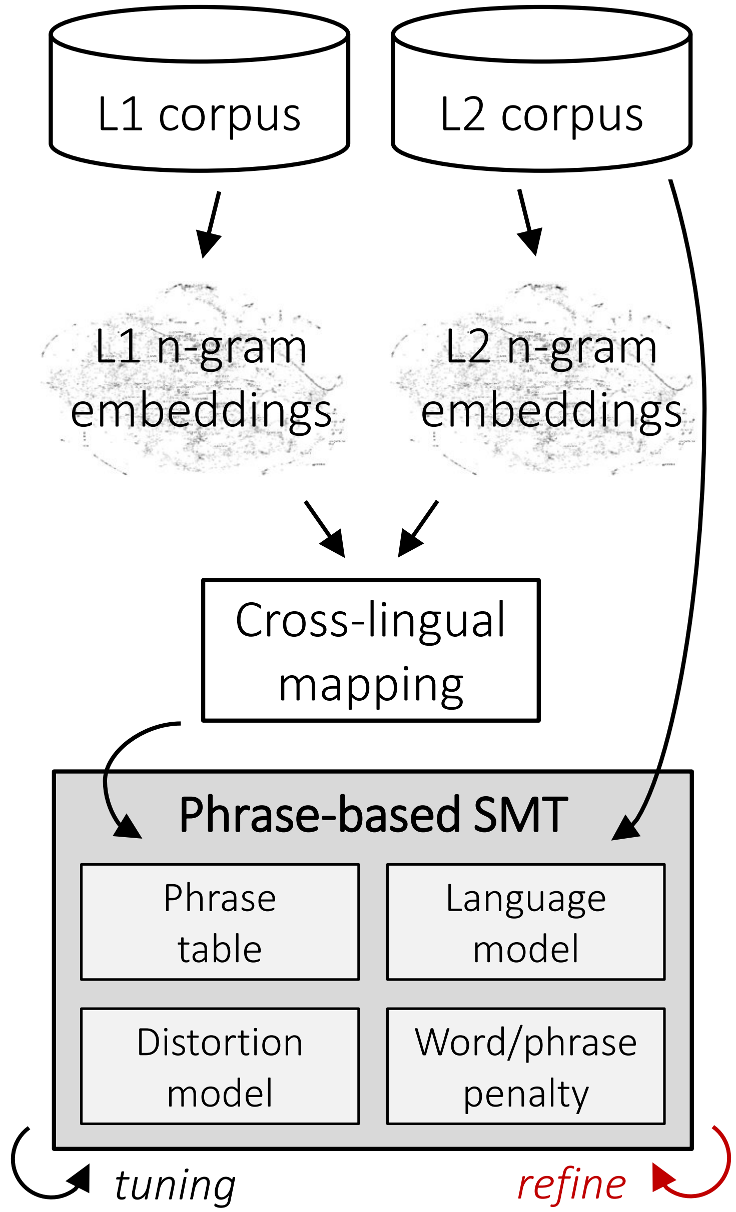


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

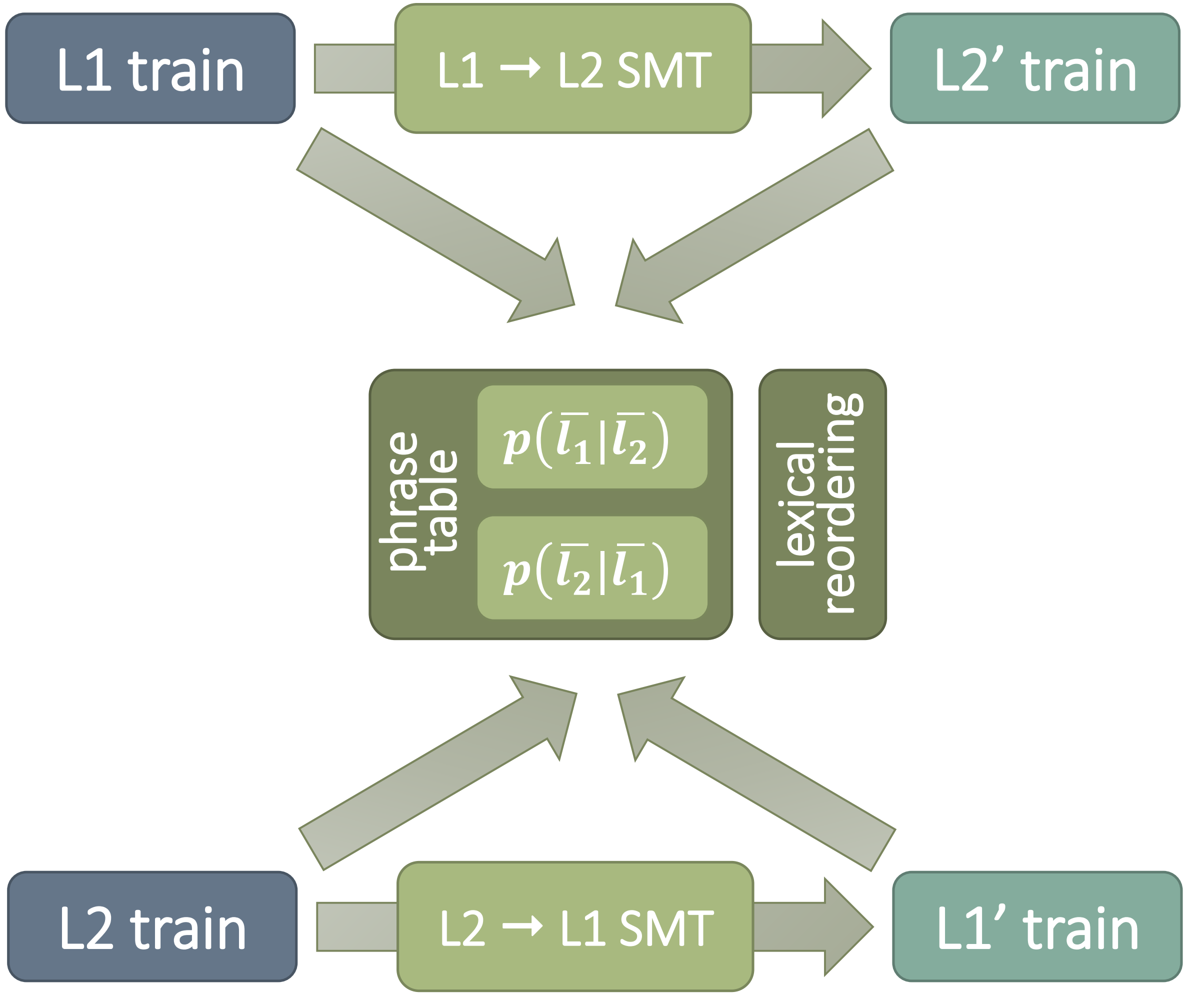


L1 train
L1 → L2 SMT
L2' train
phrase table
$p(\bar{l_1}|\bar{l_2})$
$p(\bar{l_2}|\bar{l_1})$
L2 train
L2 → L1 SMT
L1' train

Refinement

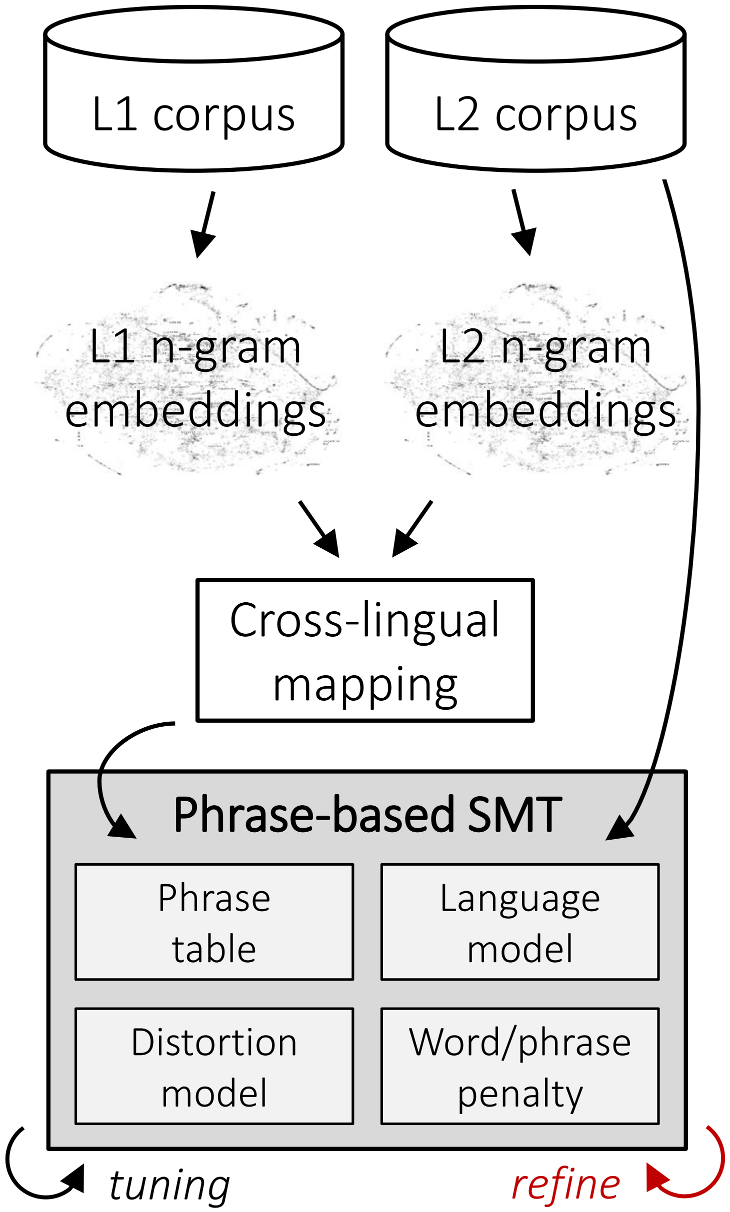


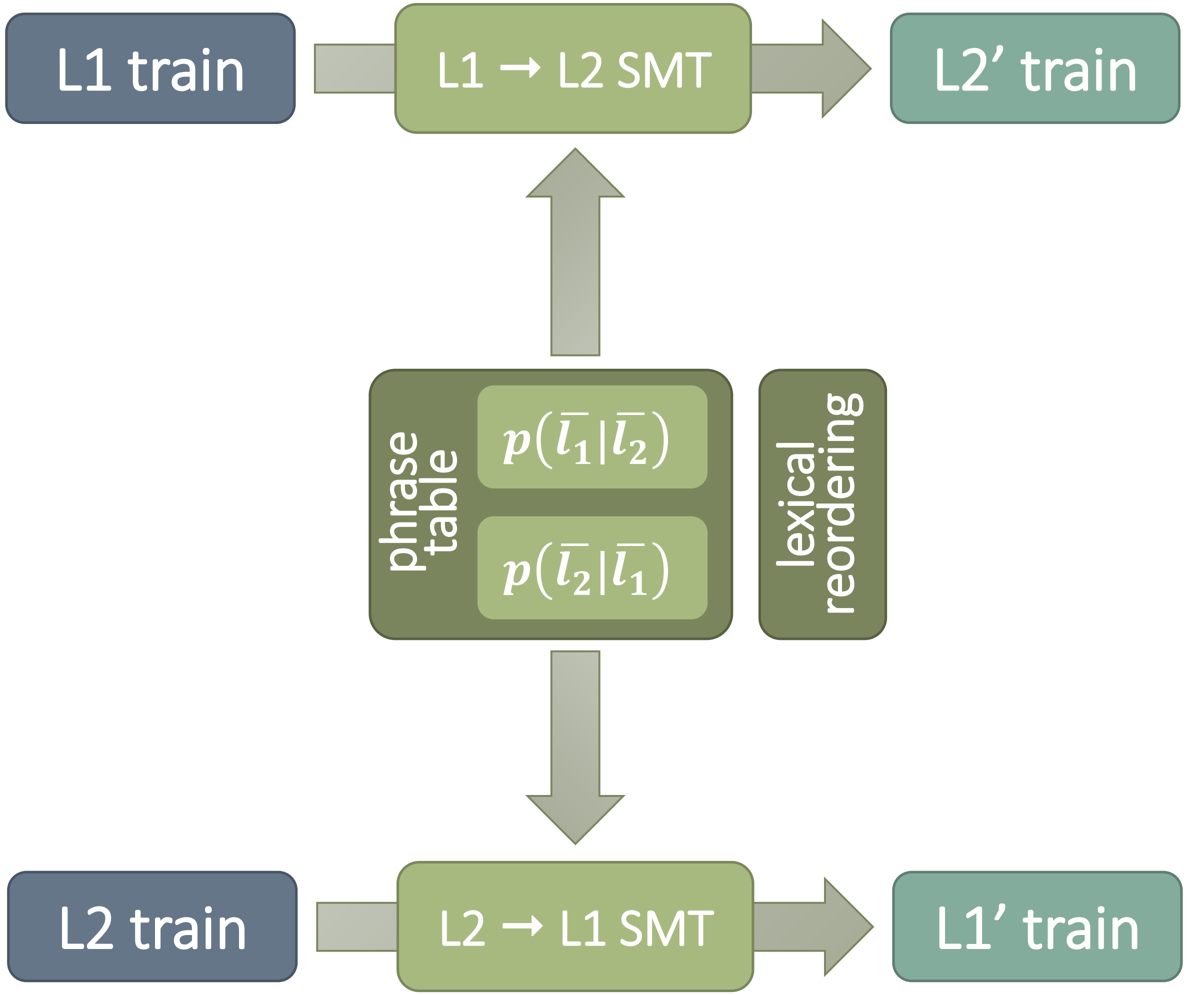
L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine



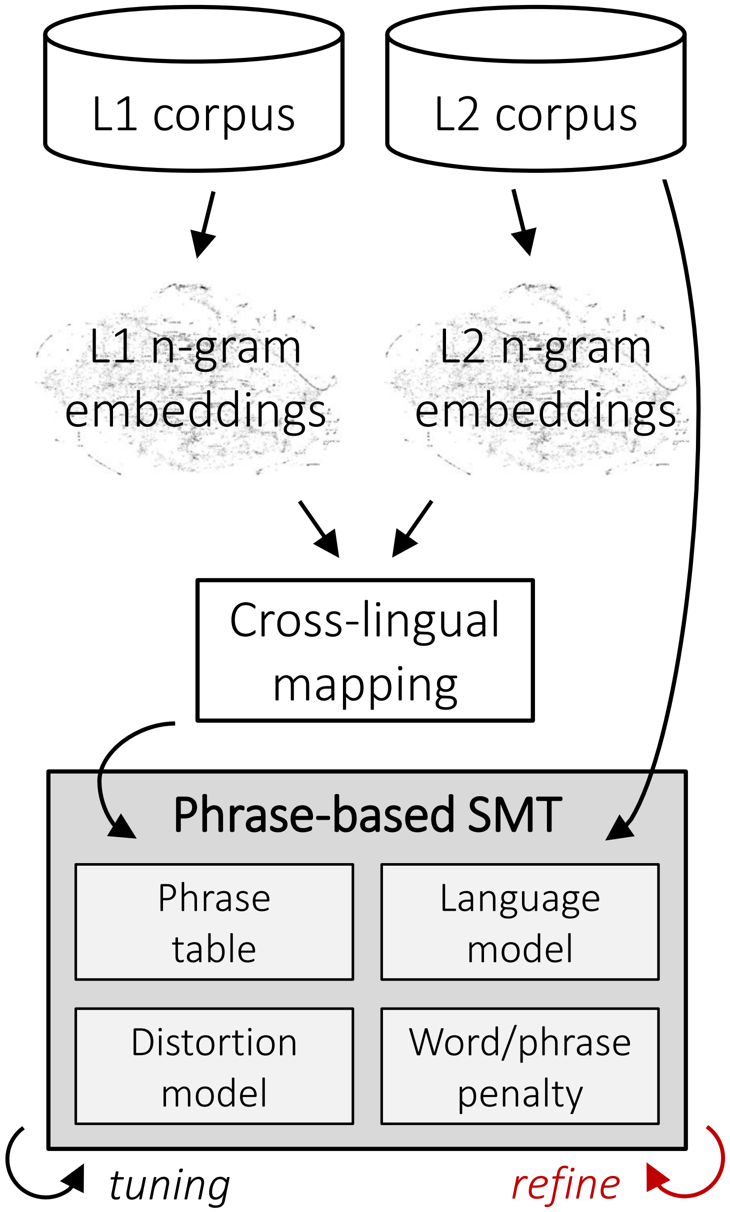
L1 train
L1 → L2 SMT
L2' train
phrase table
$p(\bar{l}_1|\bar{l}_2)$
$p(\bar{l}_2|\bar{l}_1)$
lexical reordering
L2 train
L2 → L1 SMT
L1' train

Refinement

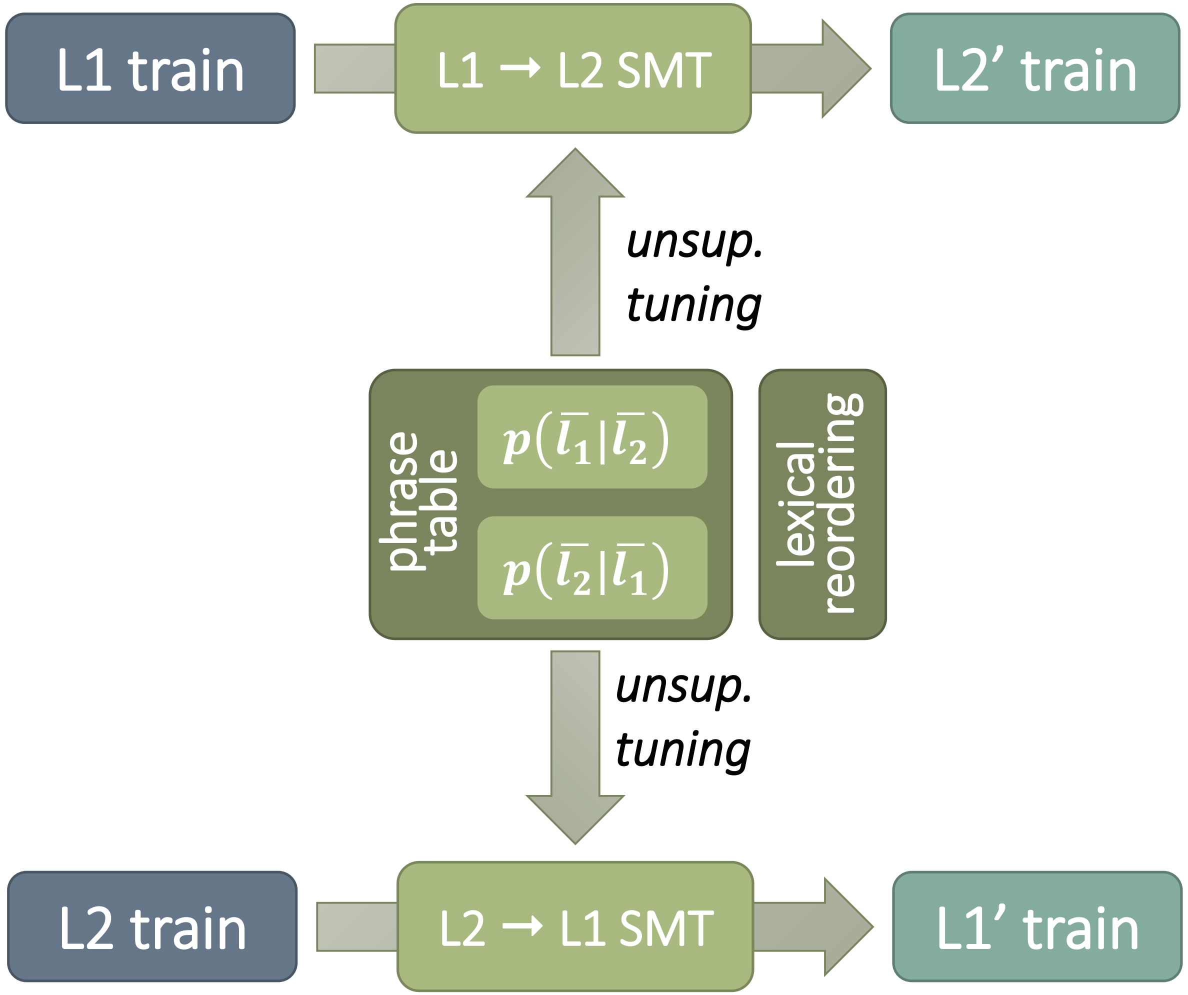

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine


L1 train
L1 → L2 SMT
L2' train
phrase table
$p(\bar{l_1}|\bar{l_2})$
$p(\bar{l_2}|\bar{l_1})$
lexical reordering
L2 train
L2 → L1 SMT
L1' train

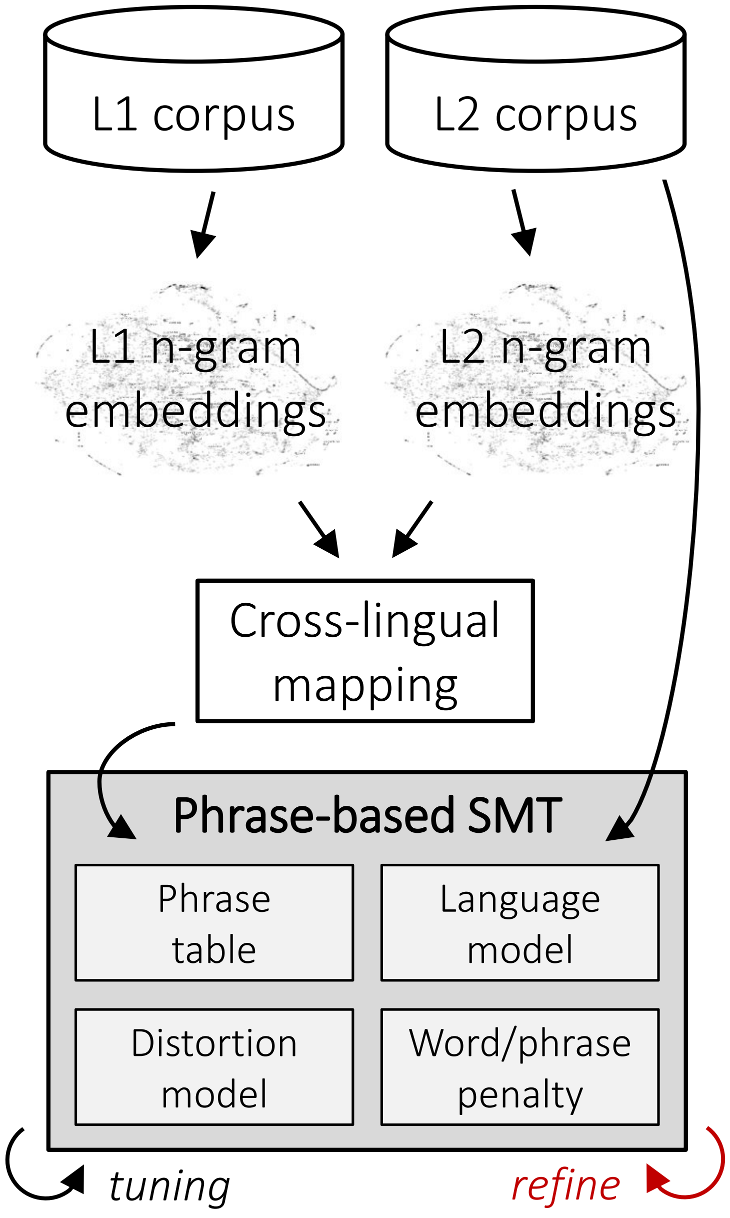
Refinement



L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine



L1 train
L1 → L2 SMT
L2' train
unsup. tuning
phrase table
$p(\bar{l_1}|\bar{l_2})$
$p(\bar{l_2}|\bar{l_1})$
lexical reordering
unsup. tuning
L2 train
L2 → L1 SMT
L1' train



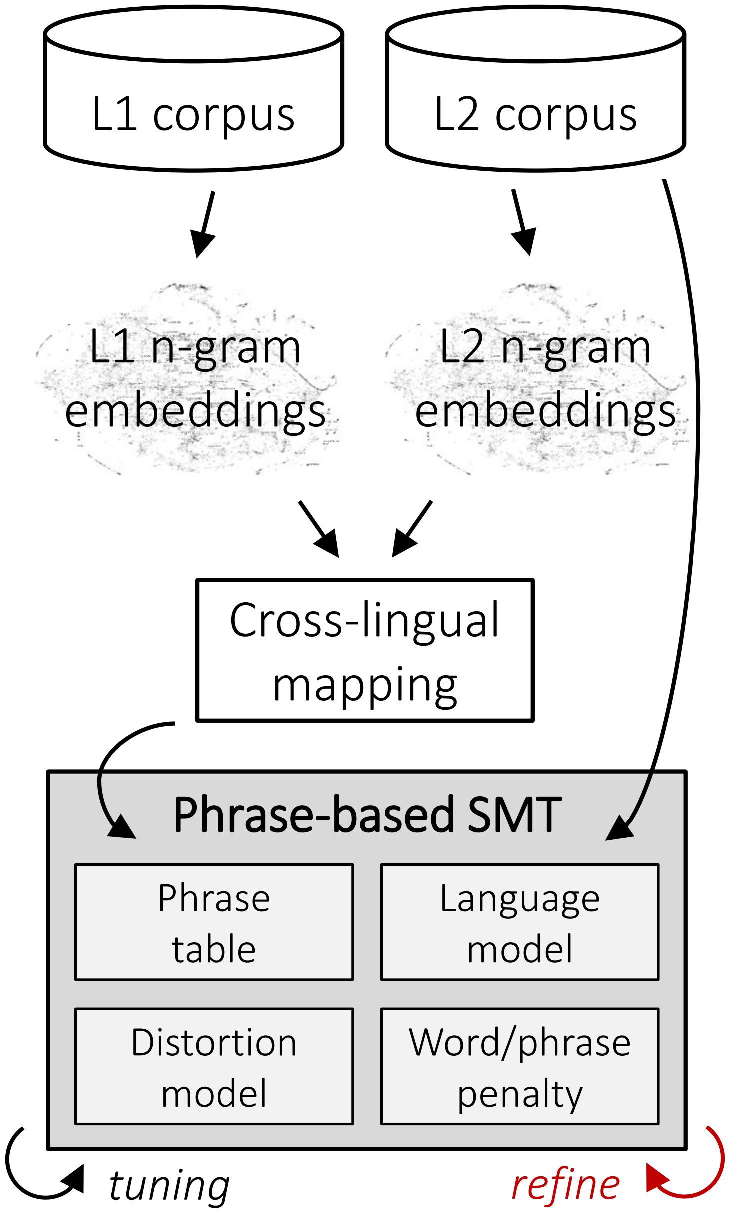
Refinement

EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

	FR-EN	EN-FR	DE-EN	EN-DE
NMT (ICLR'18)*	15.6	15.1	10.2	6.6
Initial SMT (ACL'19)	22.4	19.6	15.3	11.0
+ Tuning	23.4	21.9	15.4	11.2

*Tokenized BLEU (about 1-2 points higher)



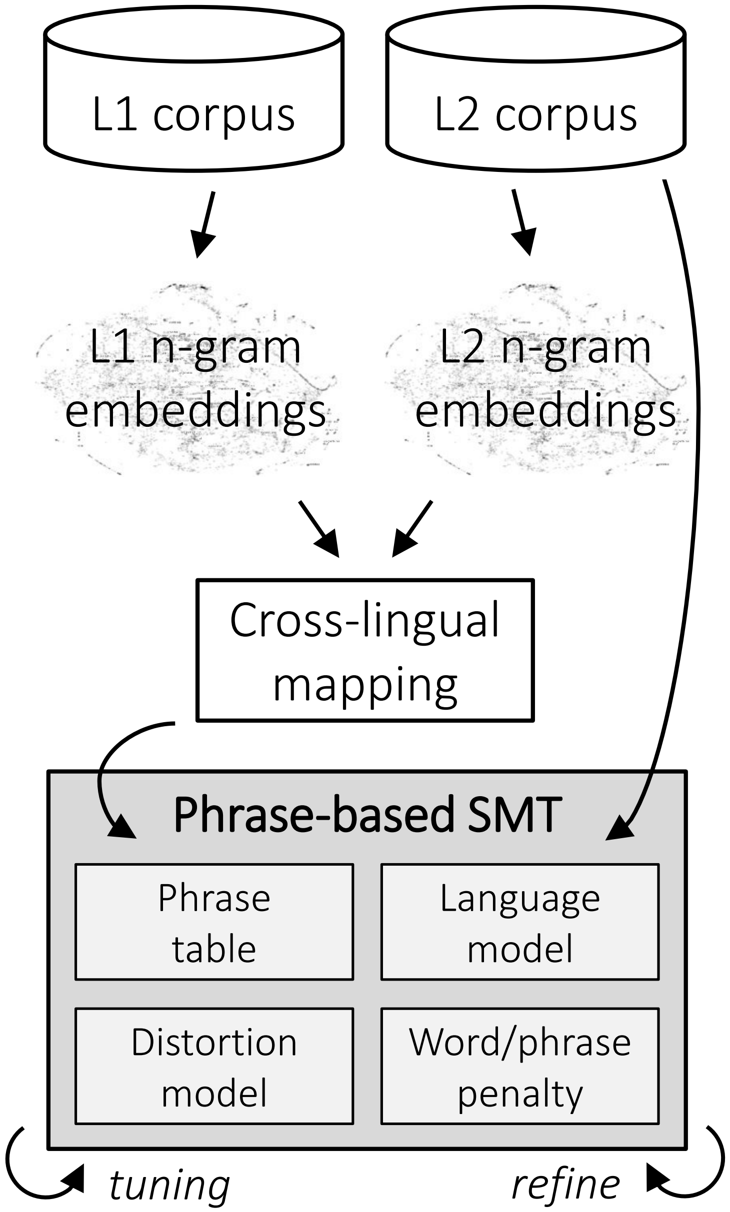
Refinement

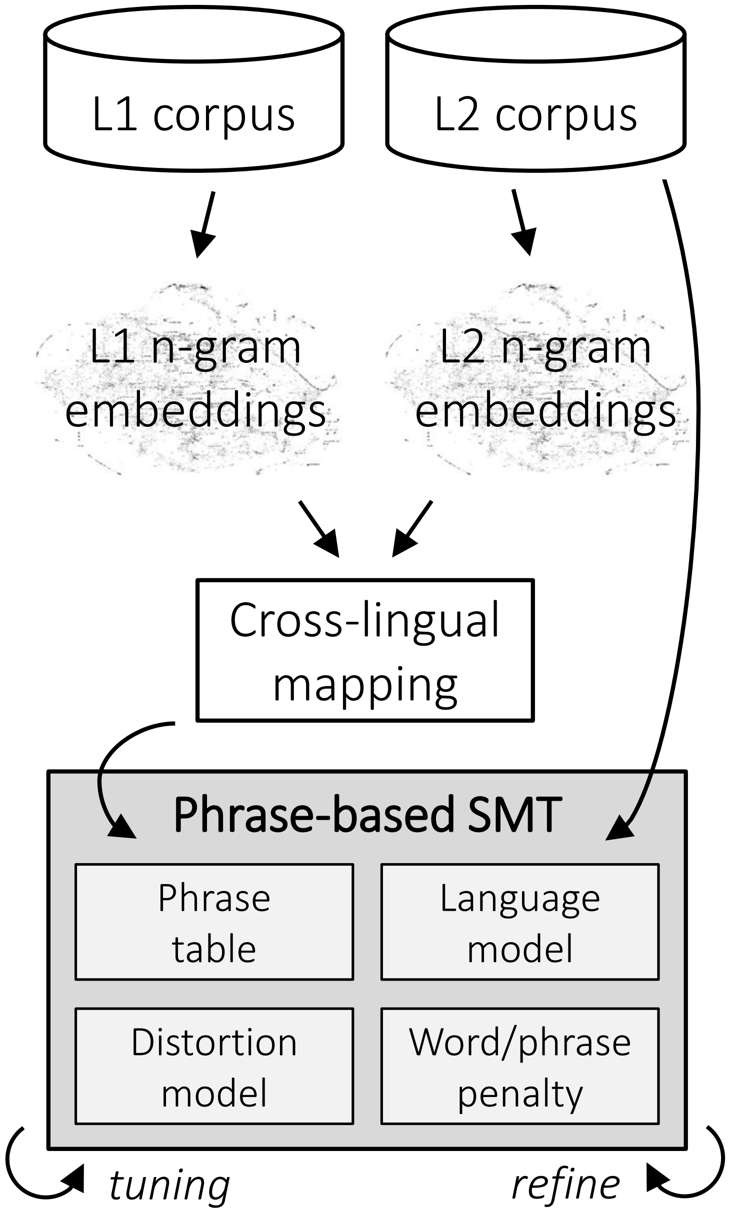
EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

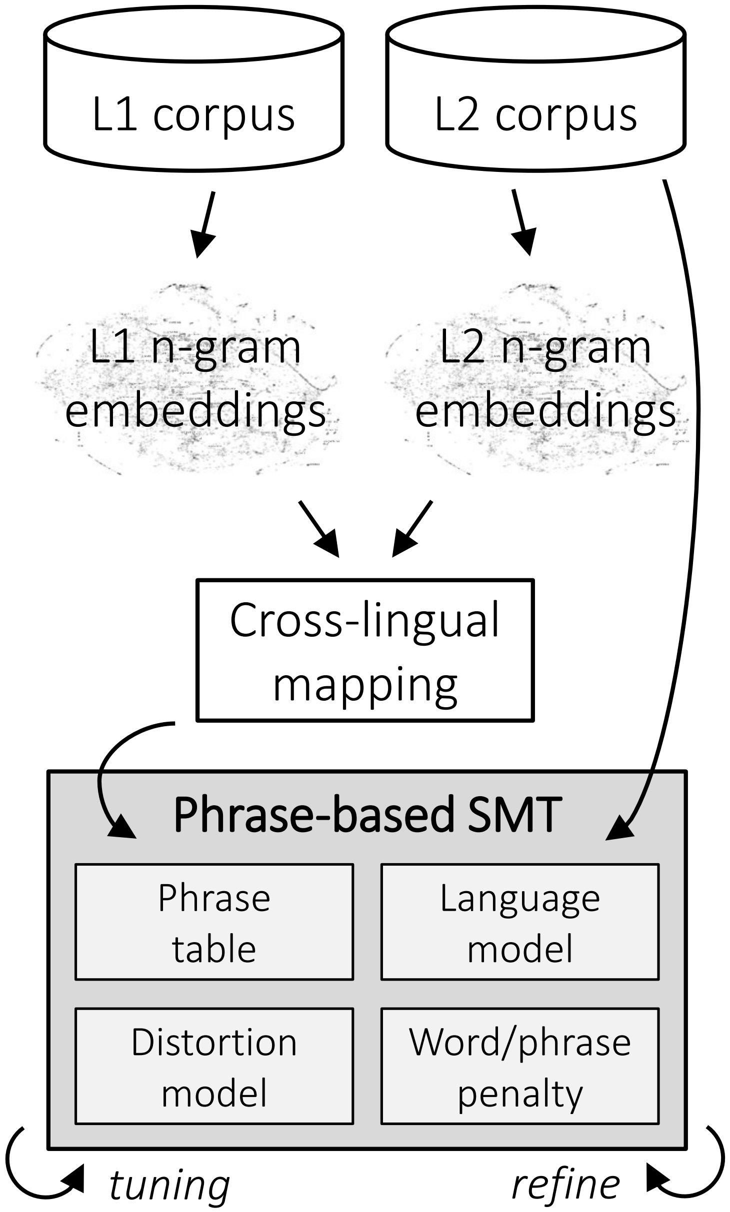
	FR-EN	EN-FR	DE-EN	EN-DE
NMT (ICLR'18)*	15.6	15.1	10.2	6.6
Initial SMT (ACL'19)	22.4	19.6	15.3	11.0
+ Tuning	23.4	21.9	15.4	11.2
+ Refinement	27.9	27.8	19.7	14.7

*Tokenized BLEU (about 1-2 points higher)


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

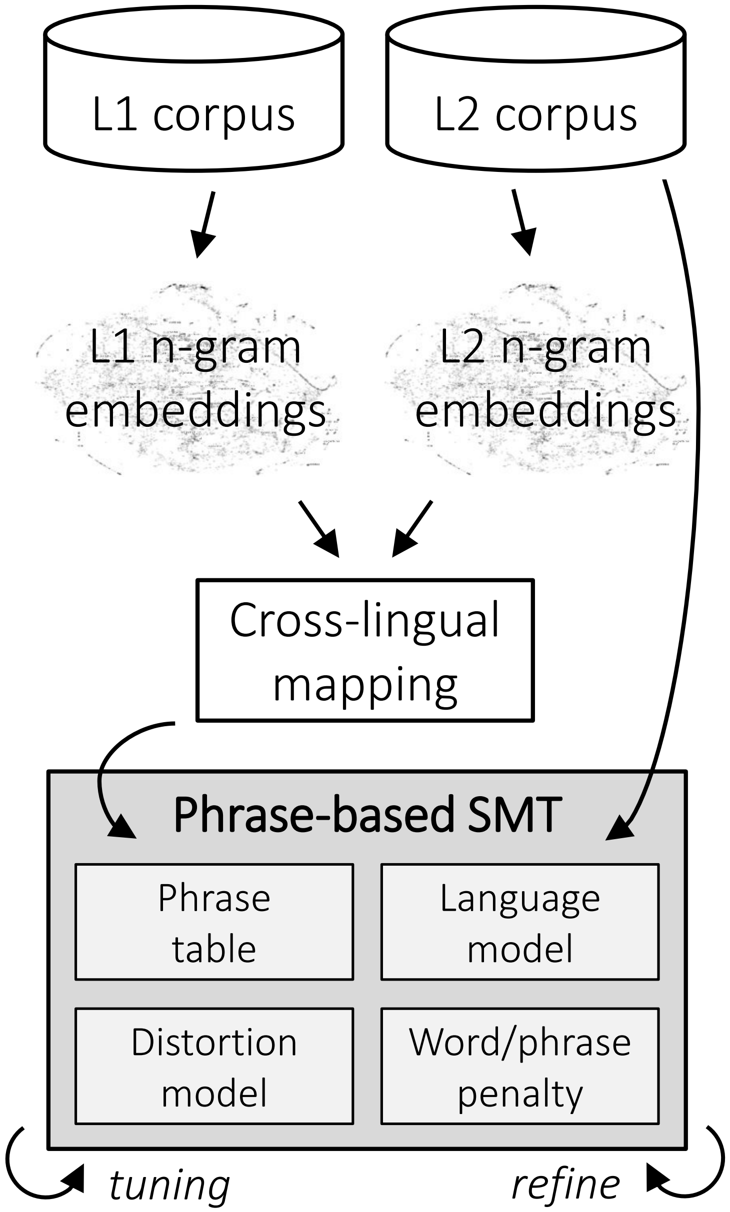

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

but…


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

but…

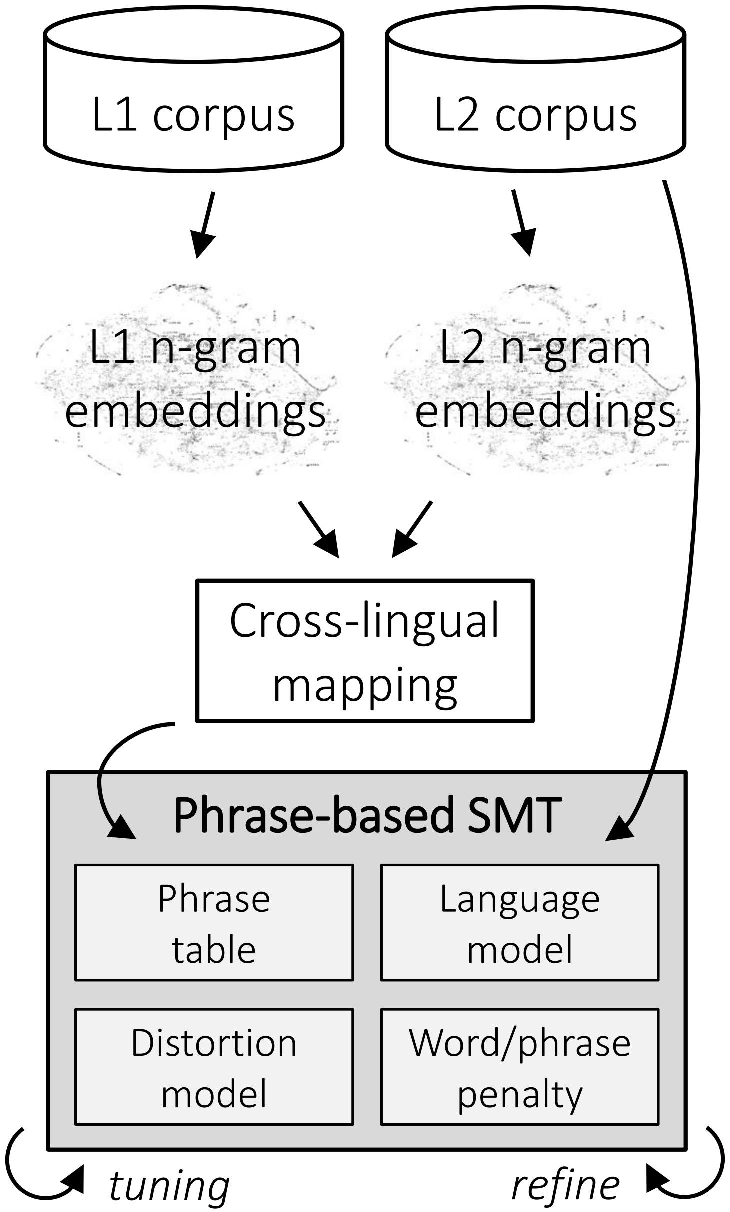
NMT >> SMT

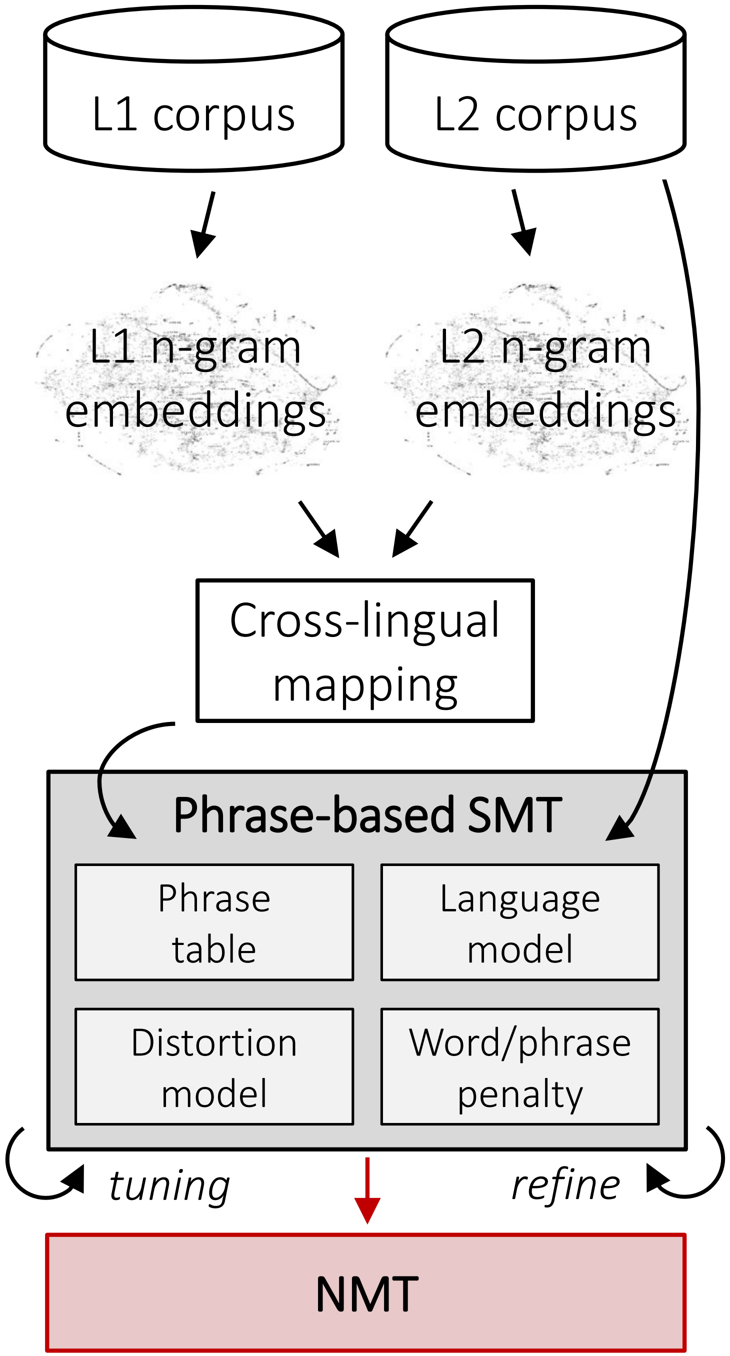


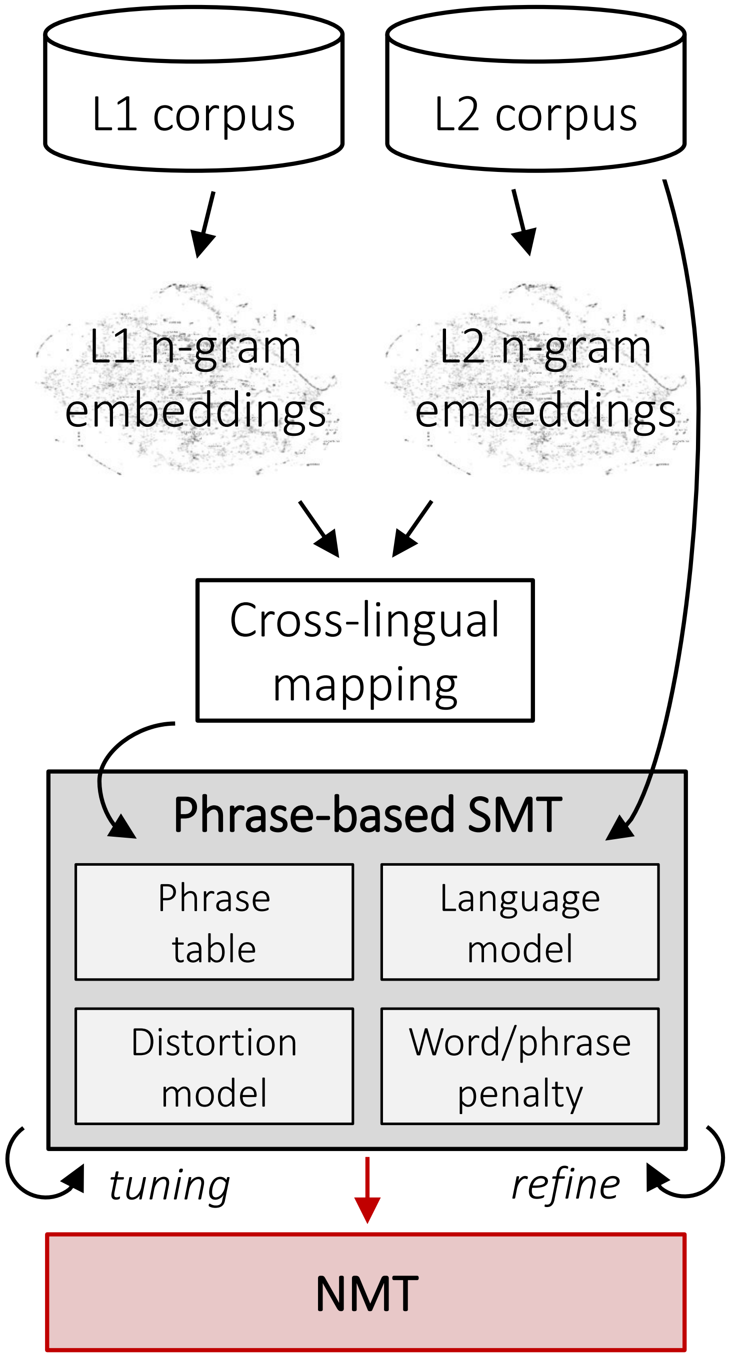
but…

NMT >> SMT

(unsupervised) SMT has a hard ceiling!

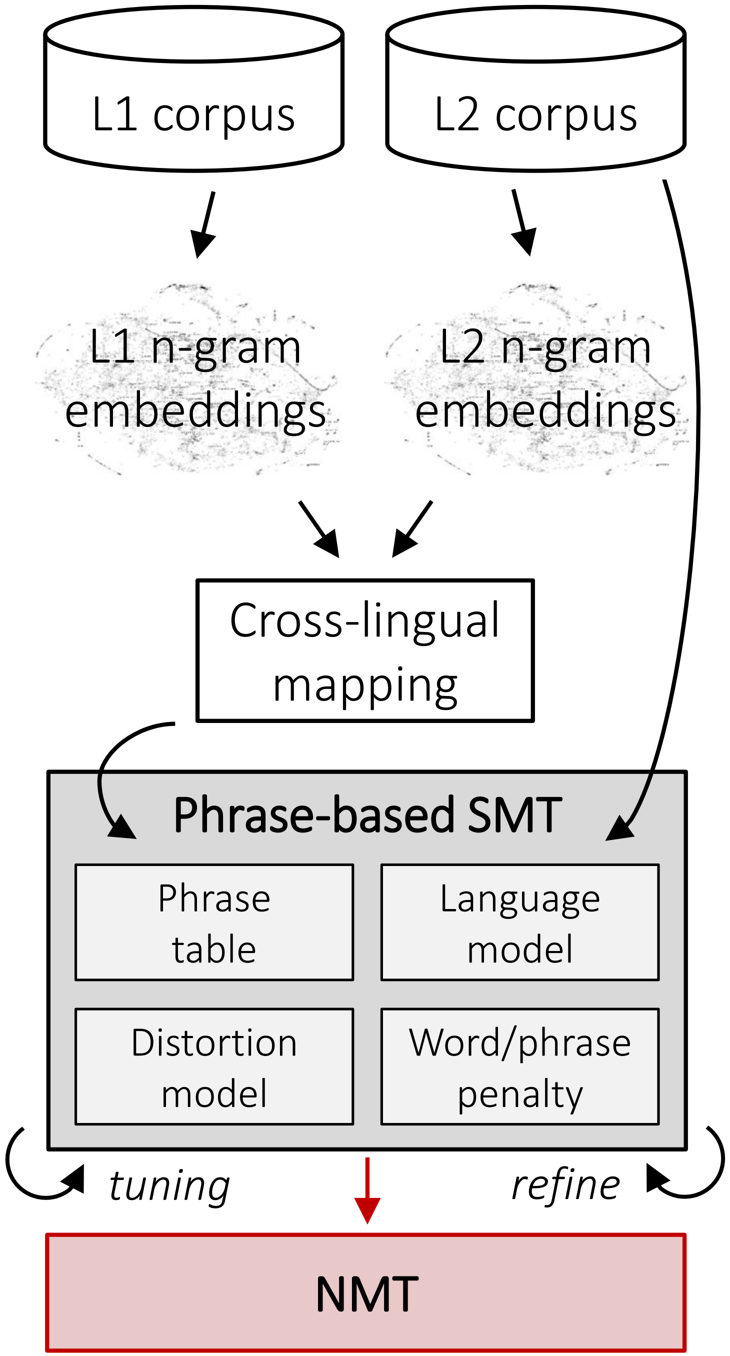

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine

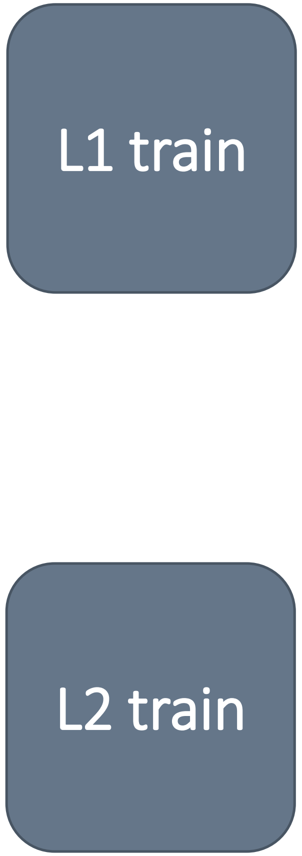

L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine
NMT



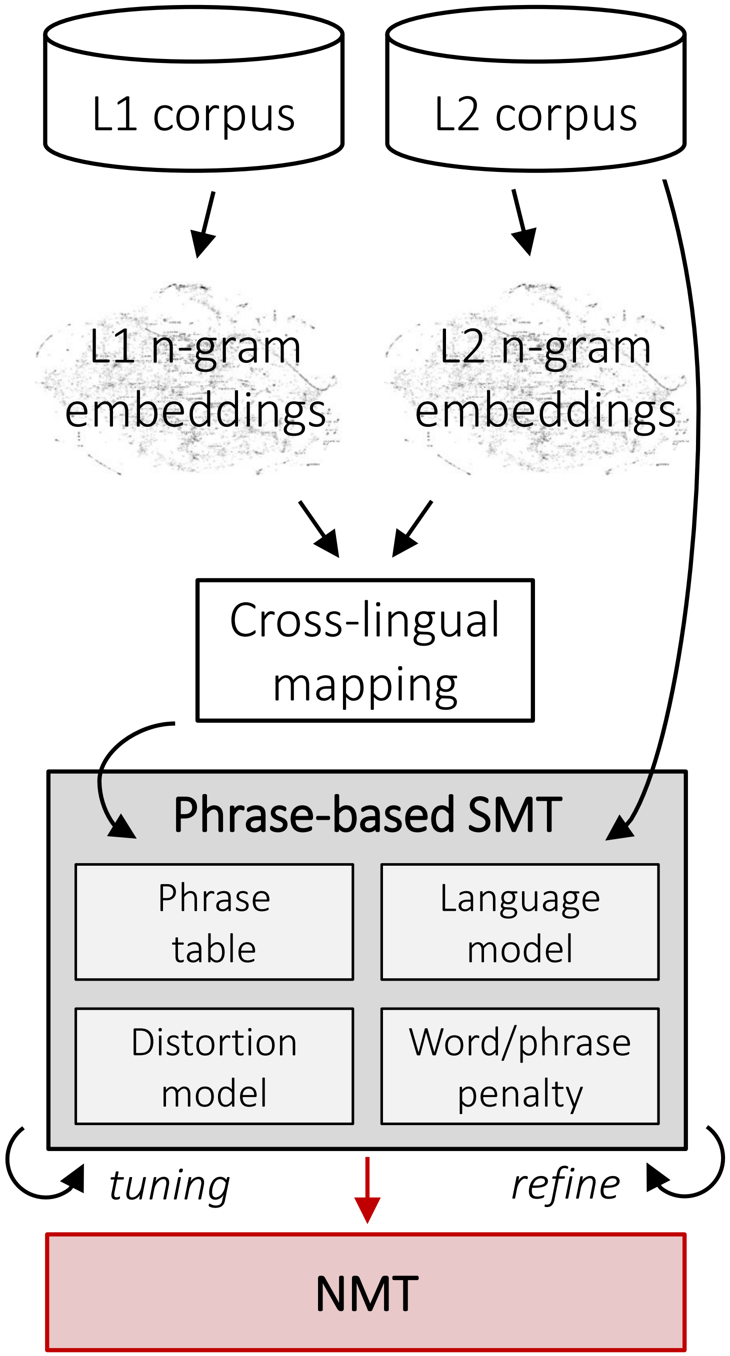
NMT hybridization

NMT hybridization

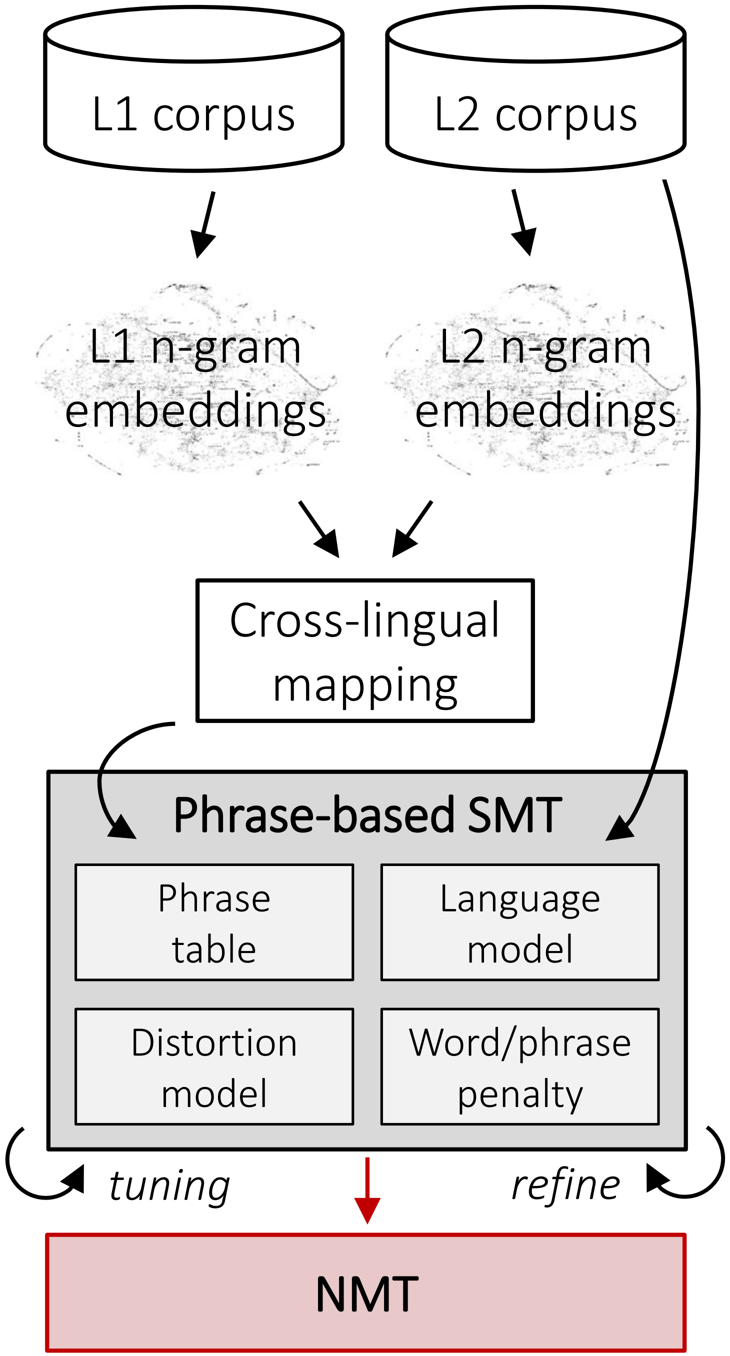


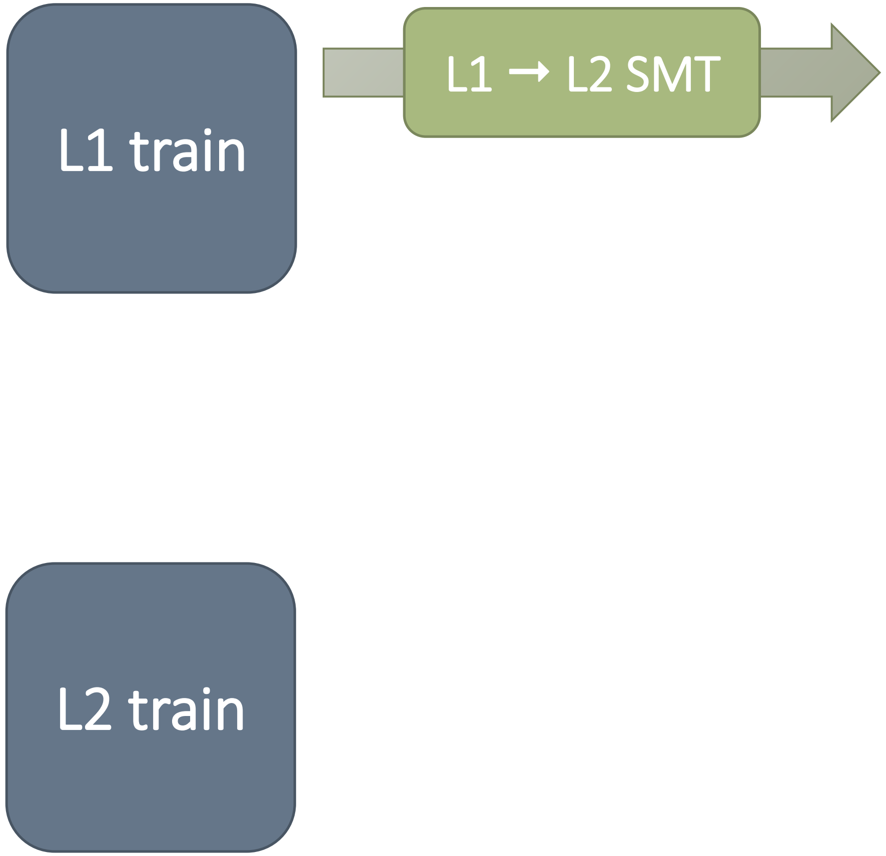


NMT hybridization

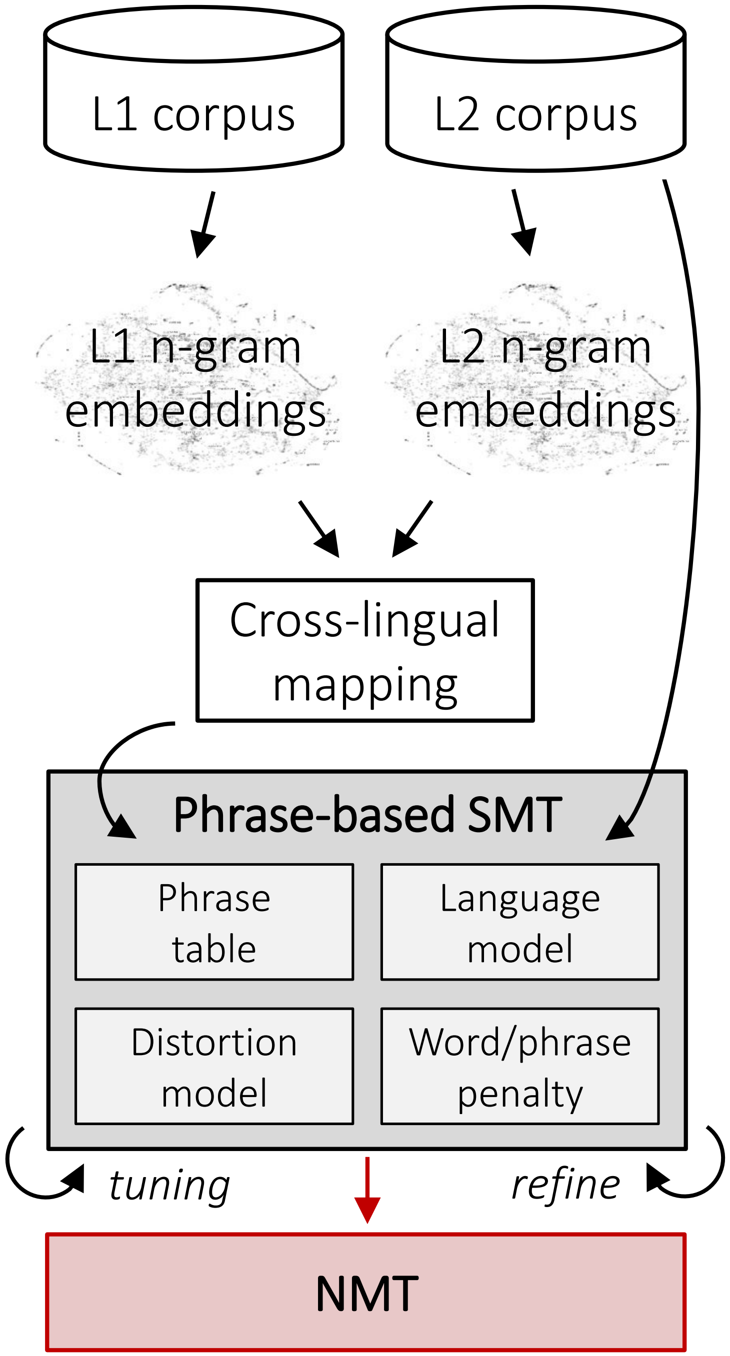


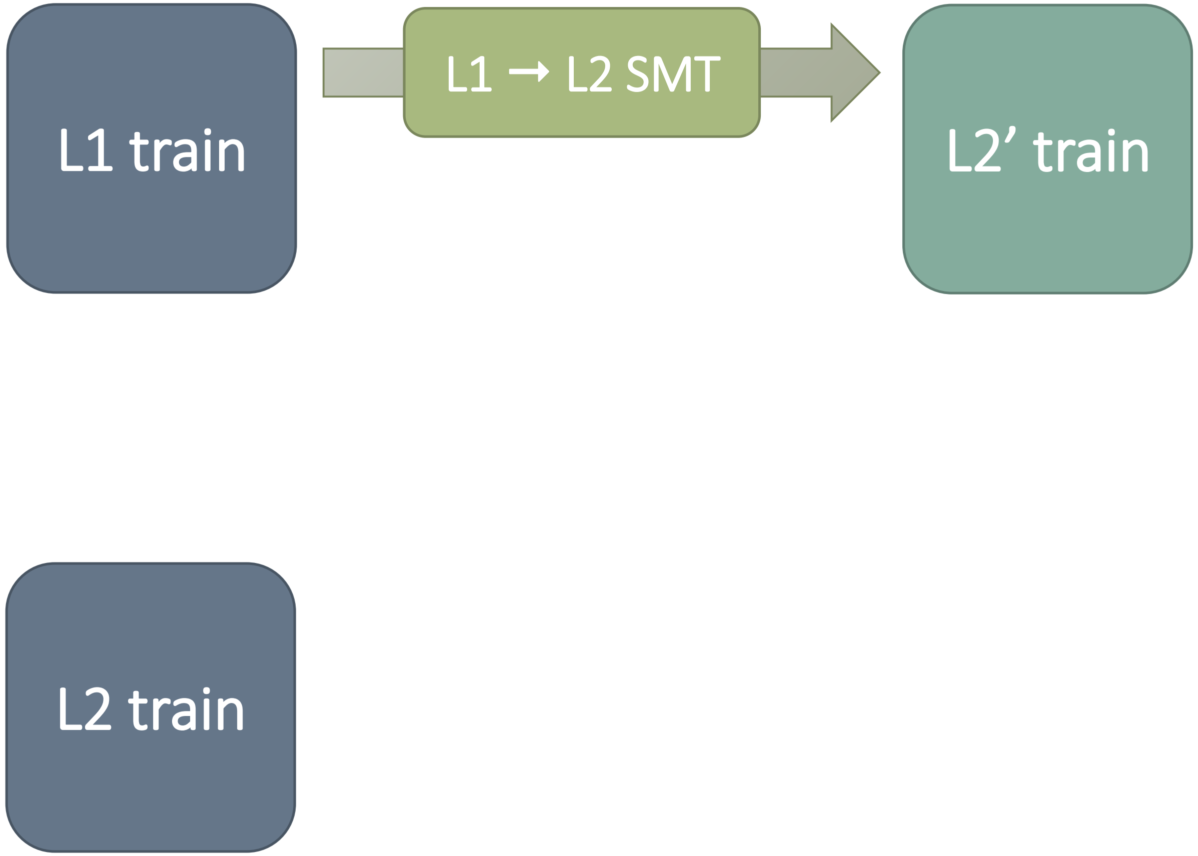
NMT hybridization



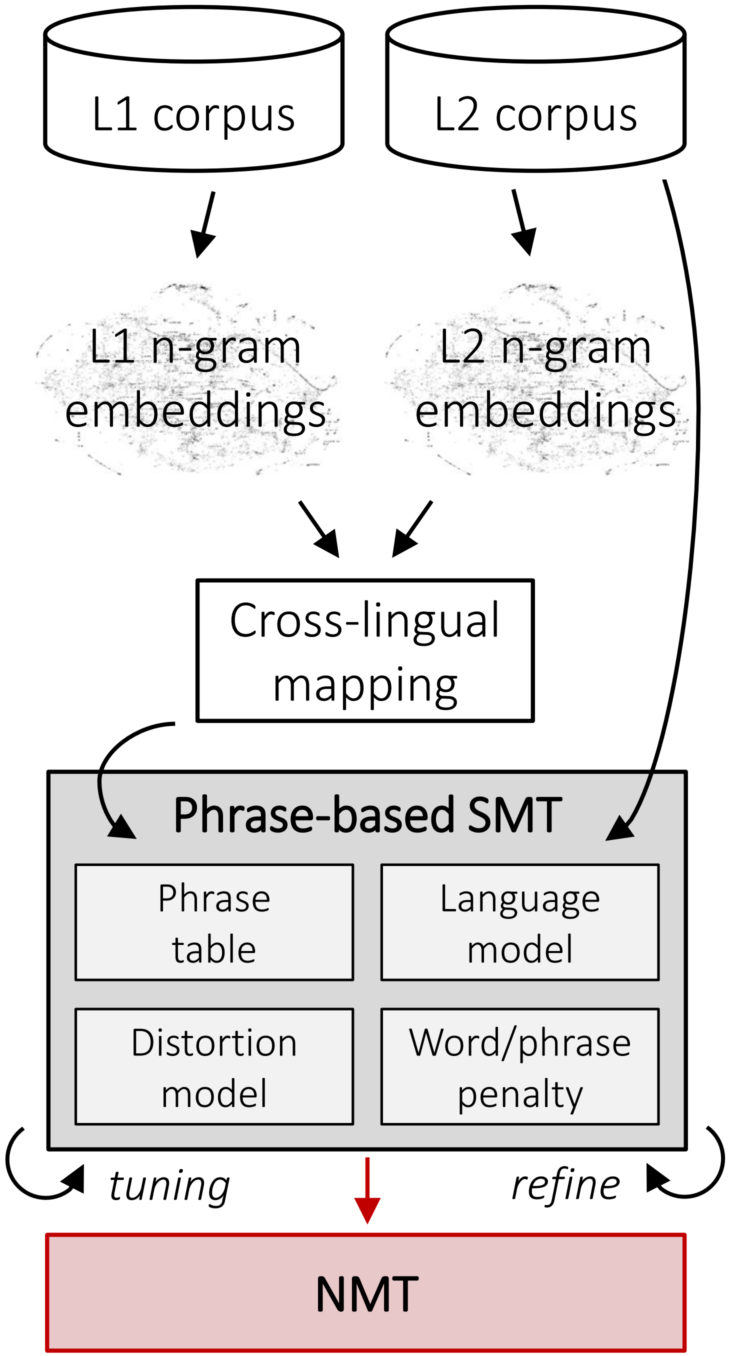


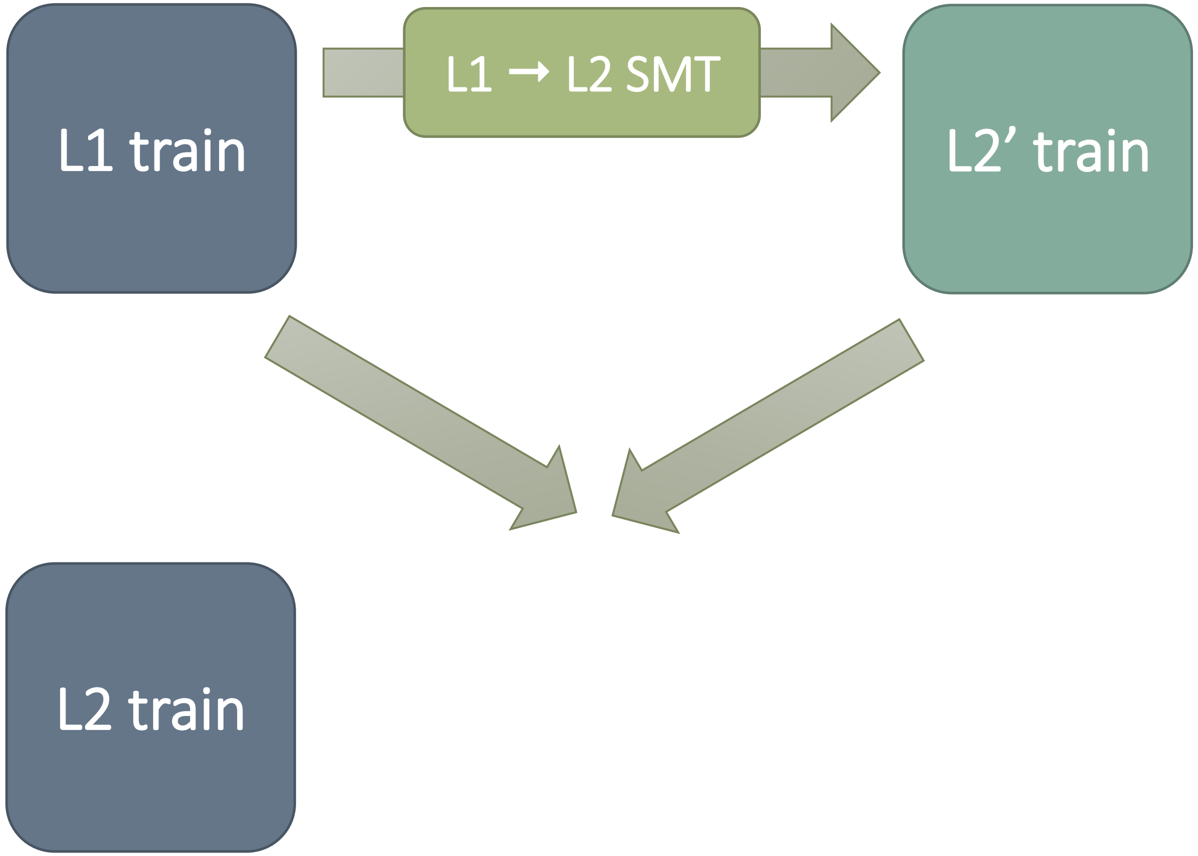
NMT hybridization



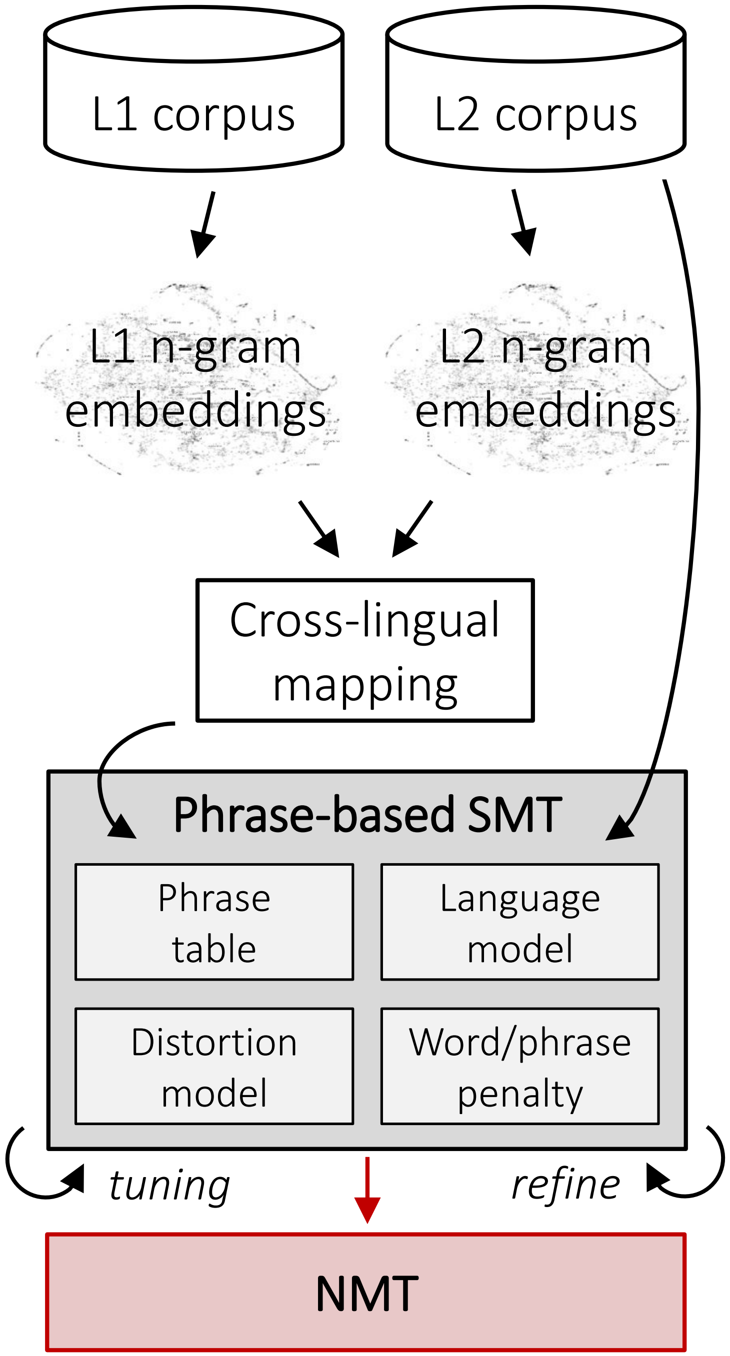


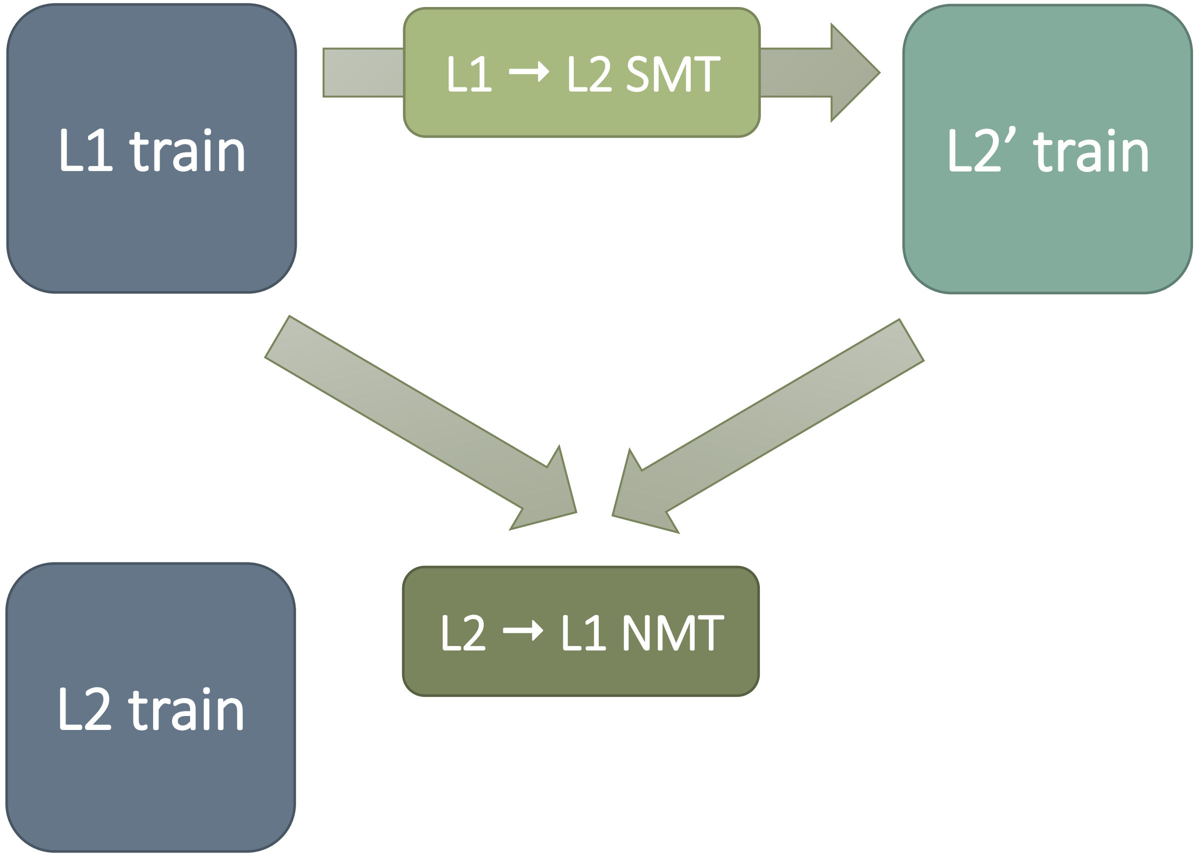
NMT hybridization

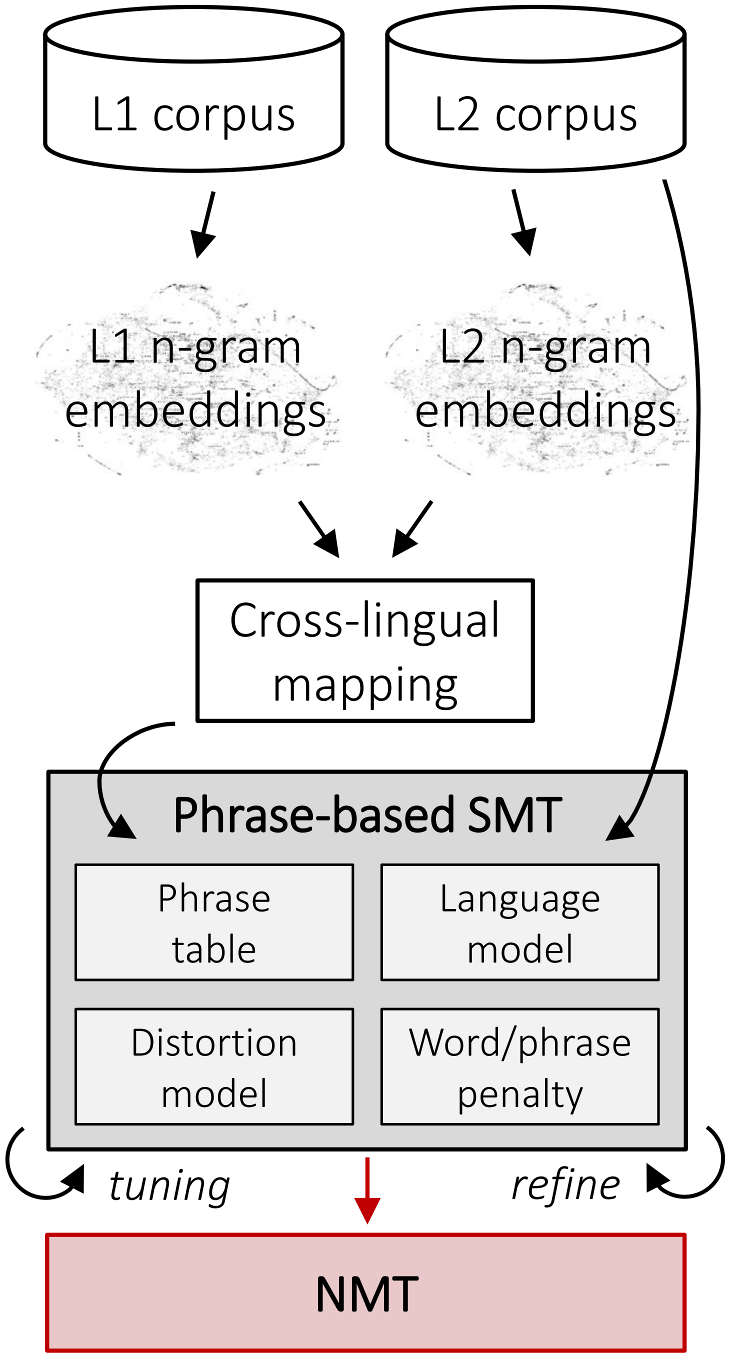




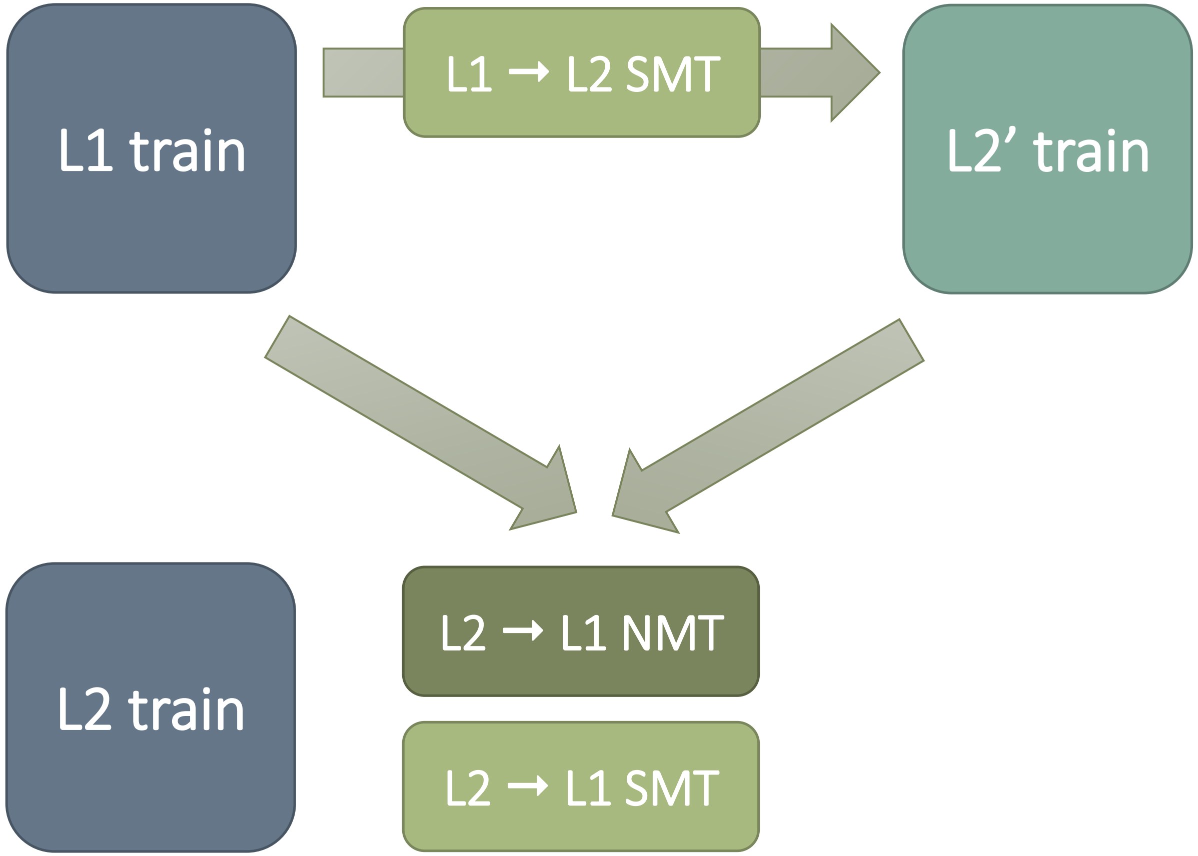
NMT hybridization

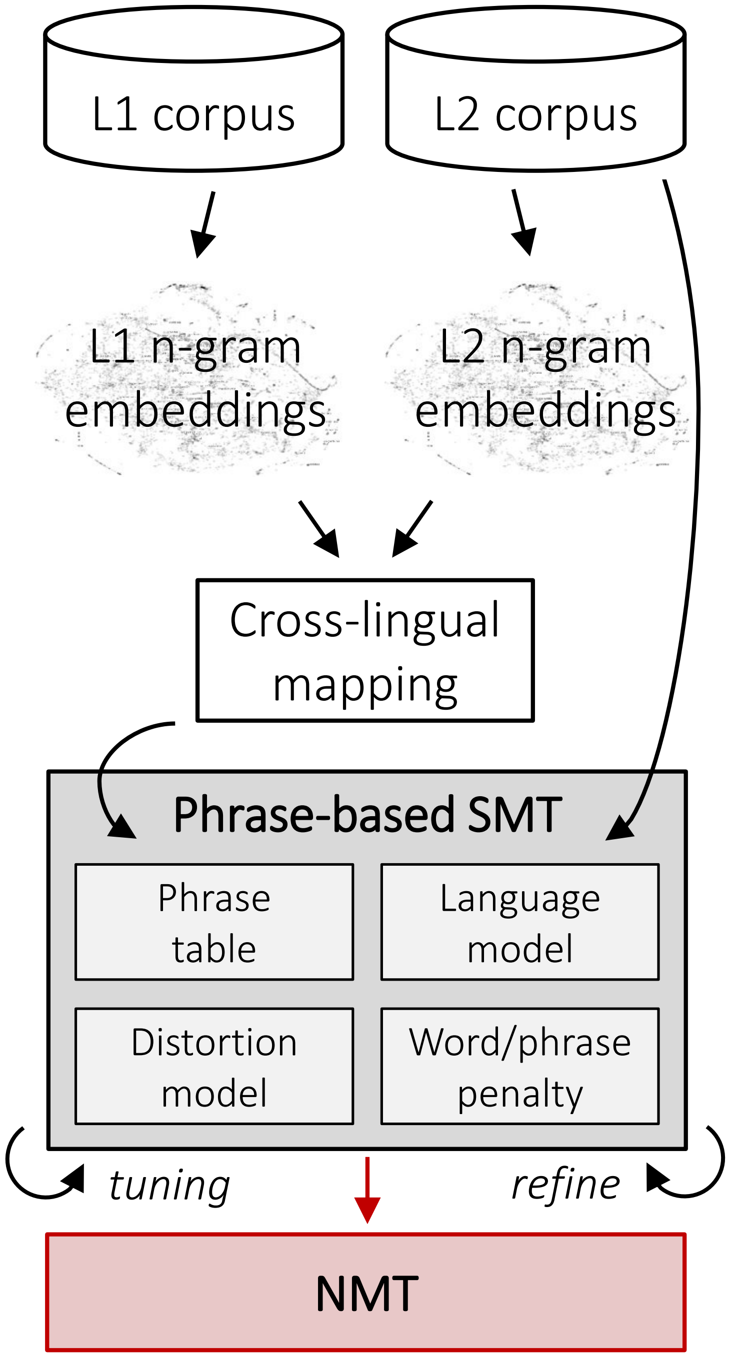




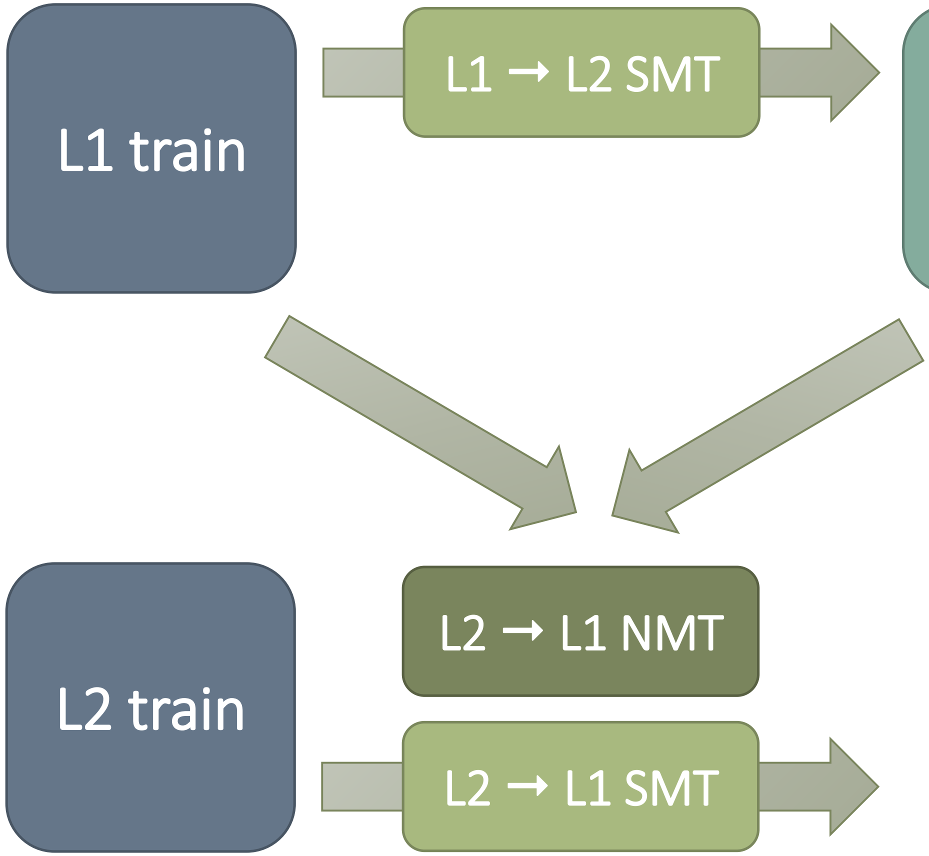


NMT hybridization

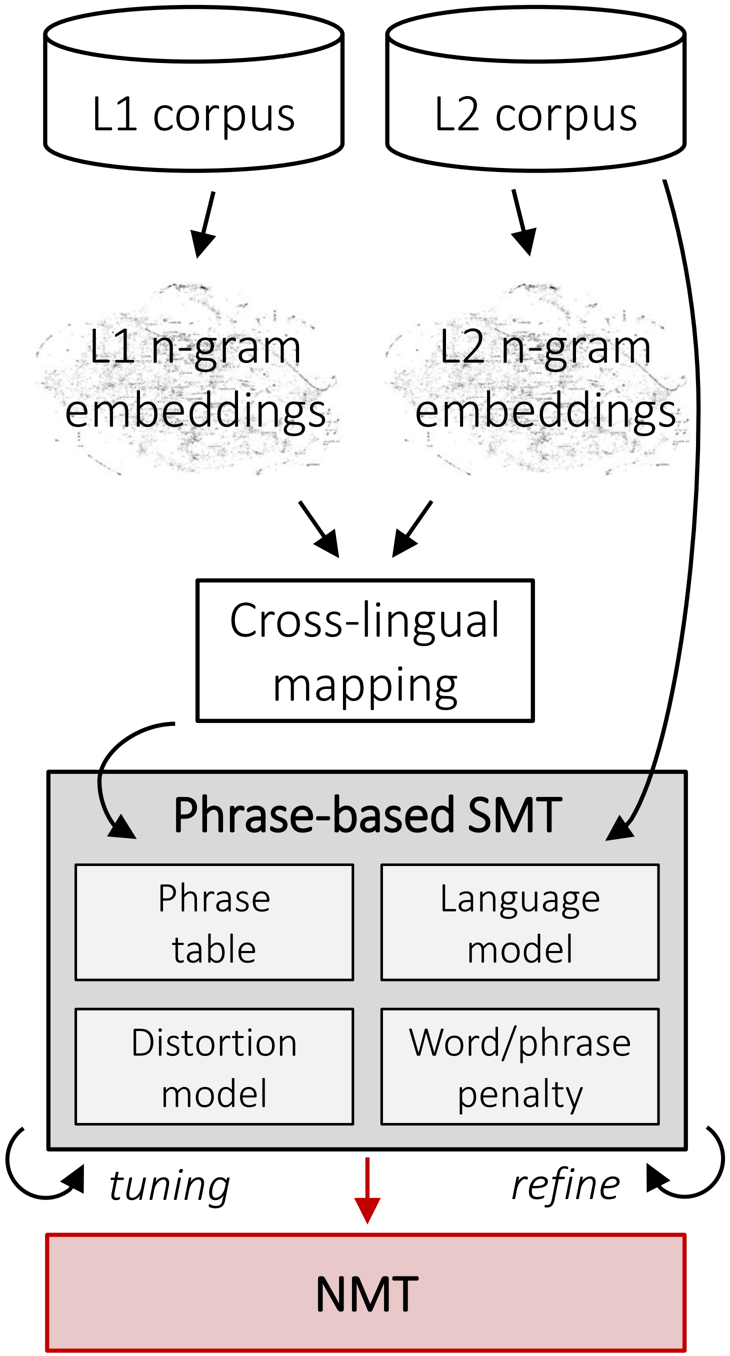


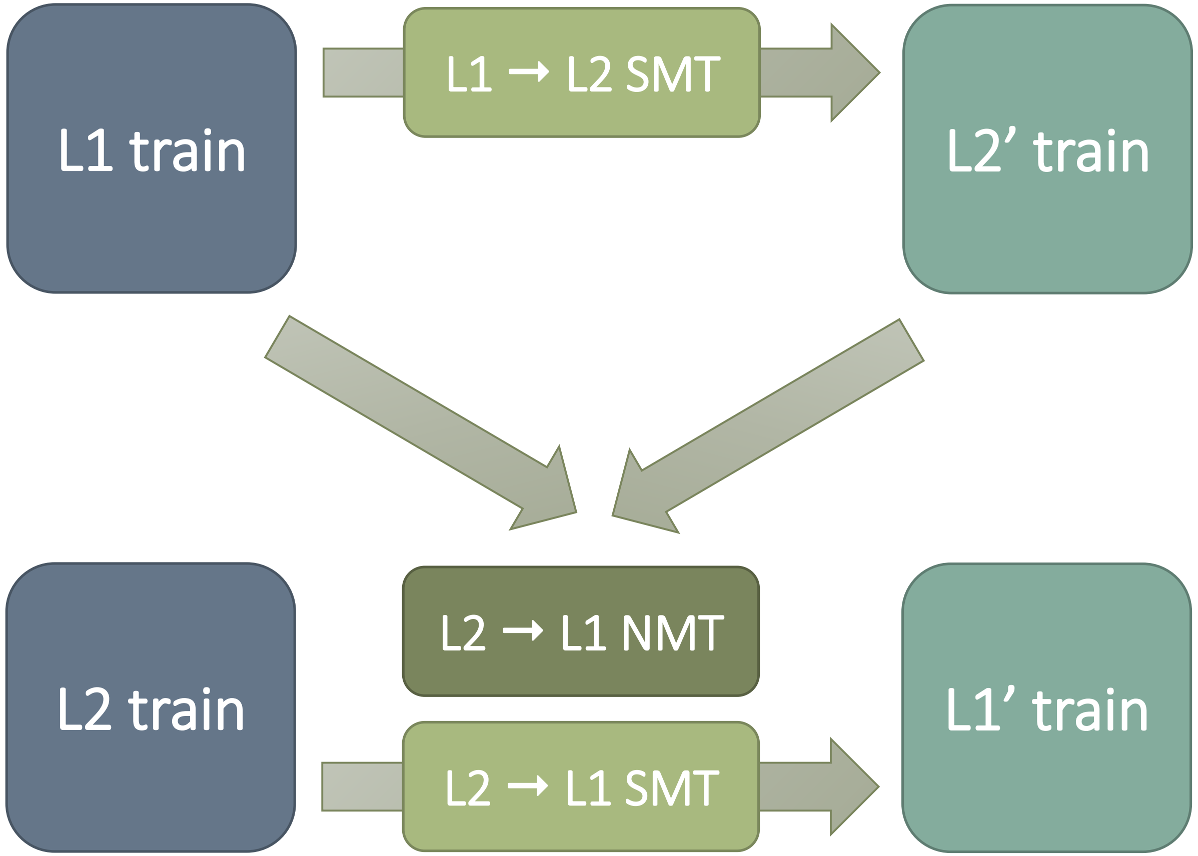


NMT hybridization

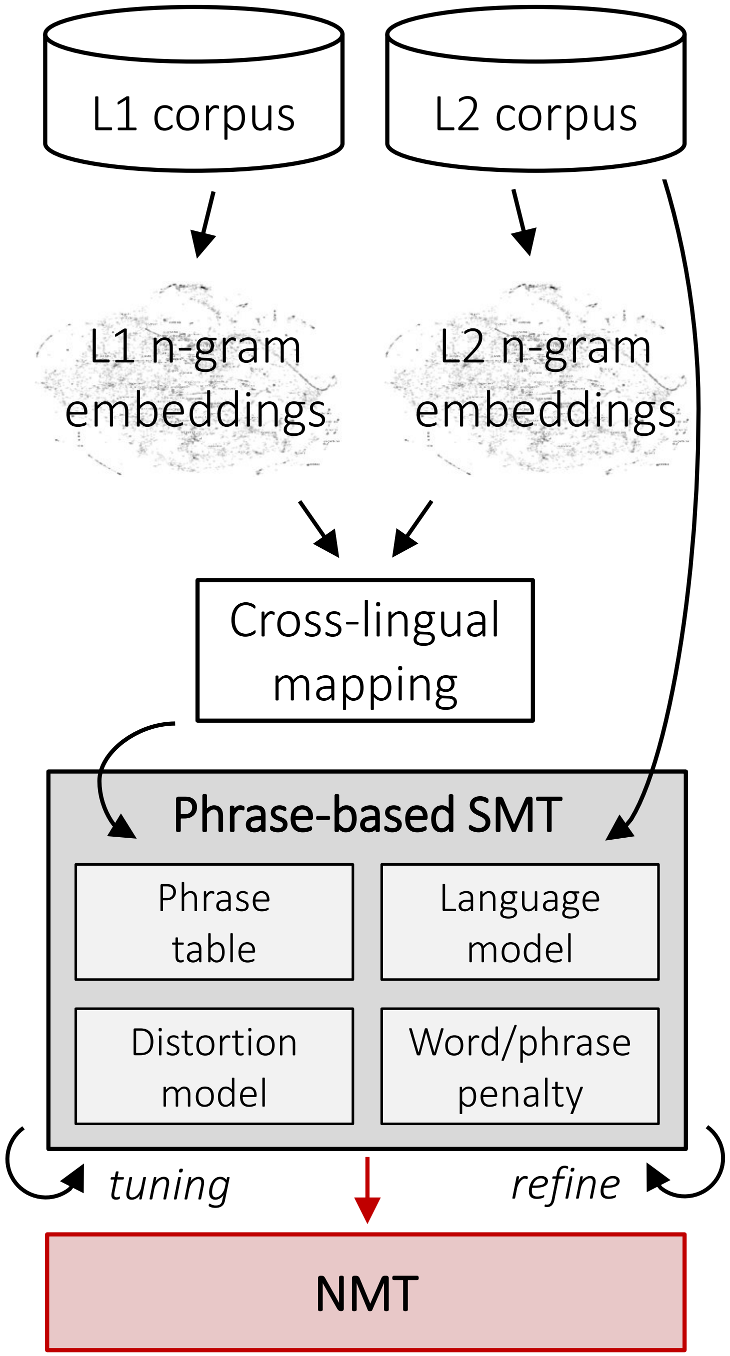


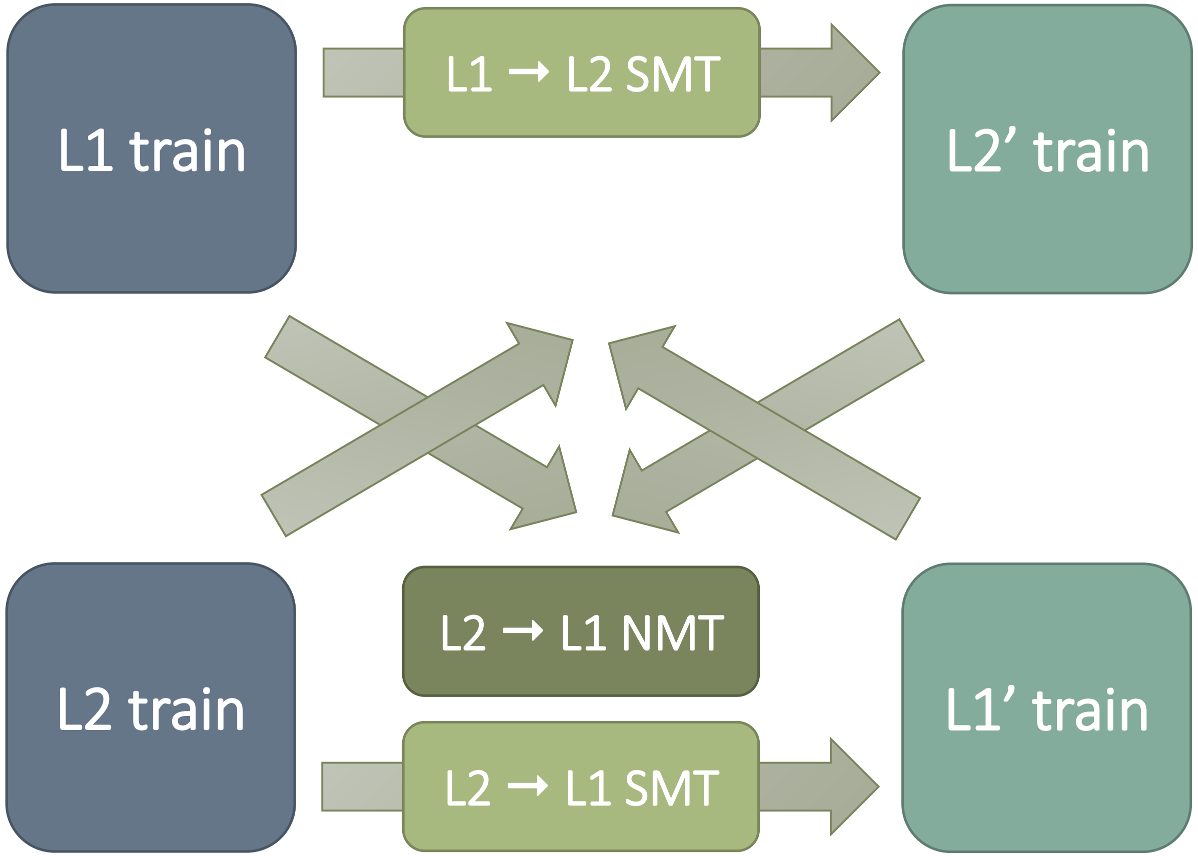
NMT hybridization



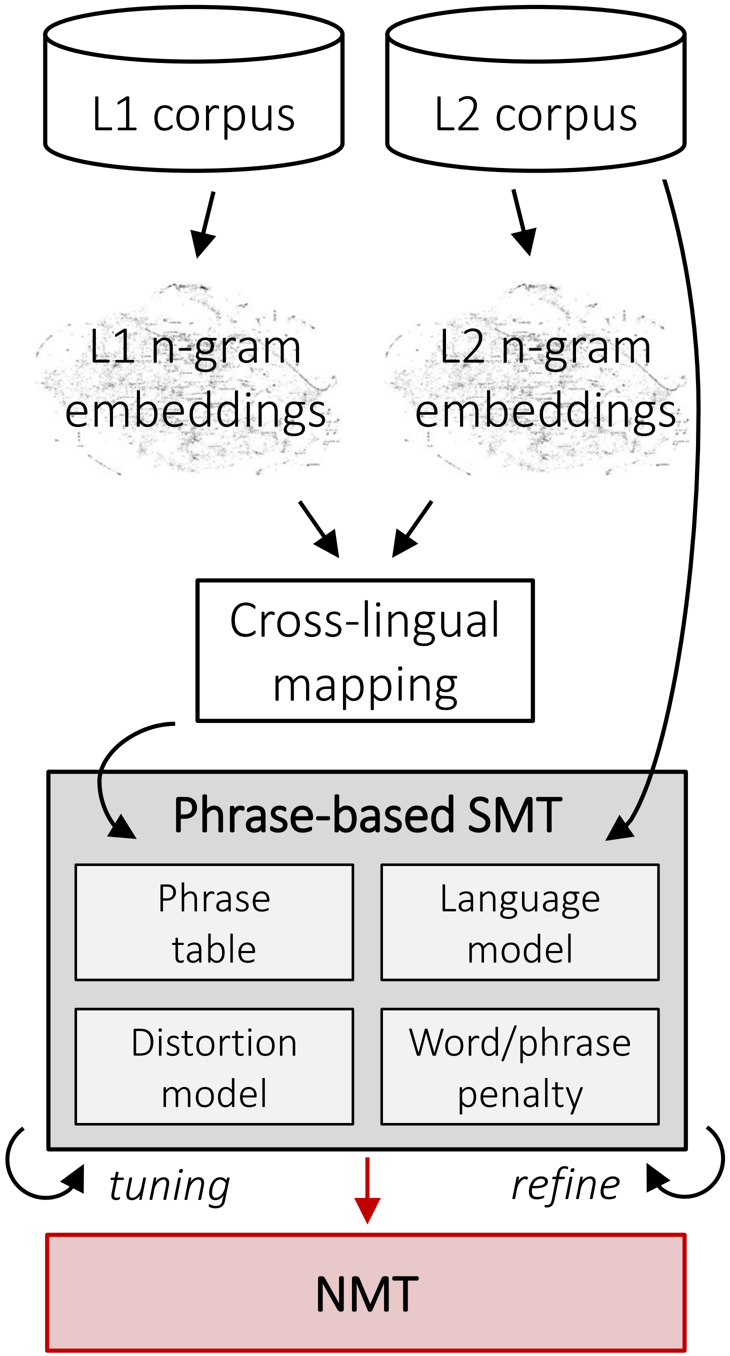


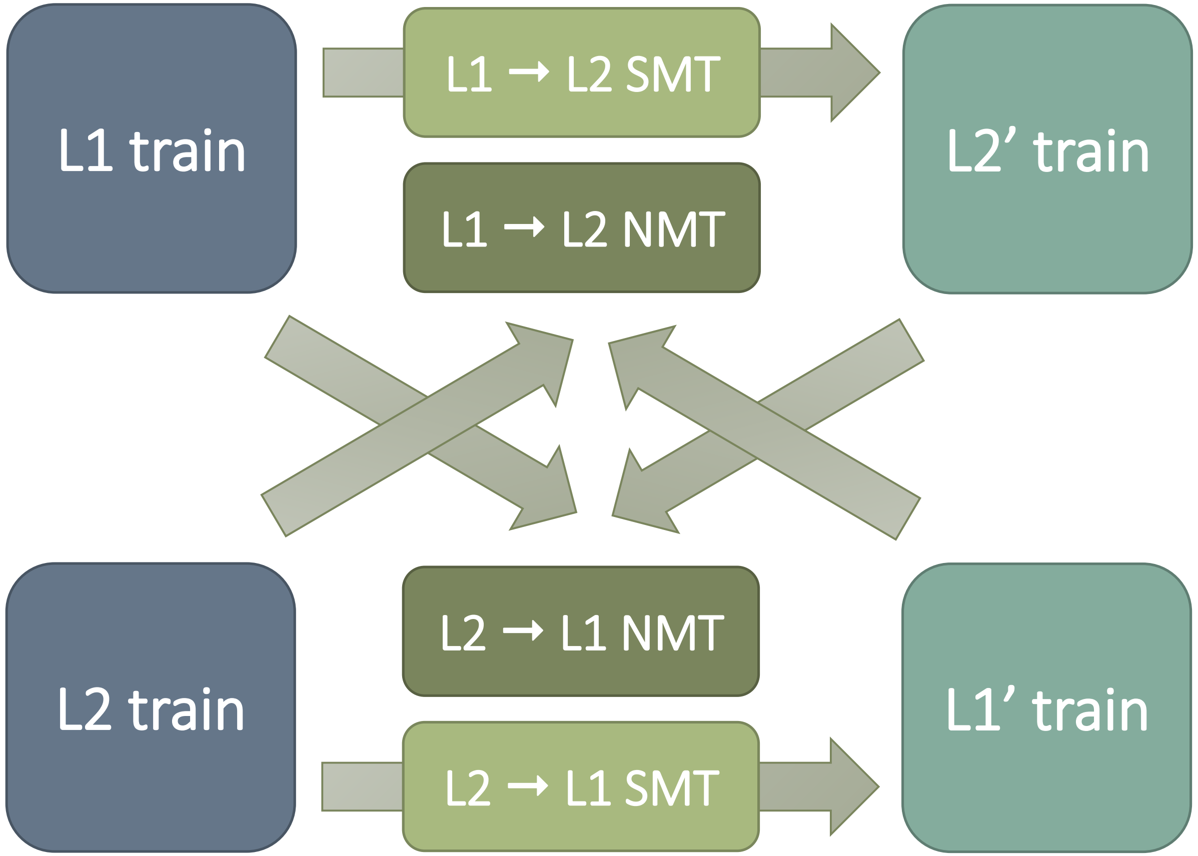
NMT hybridization



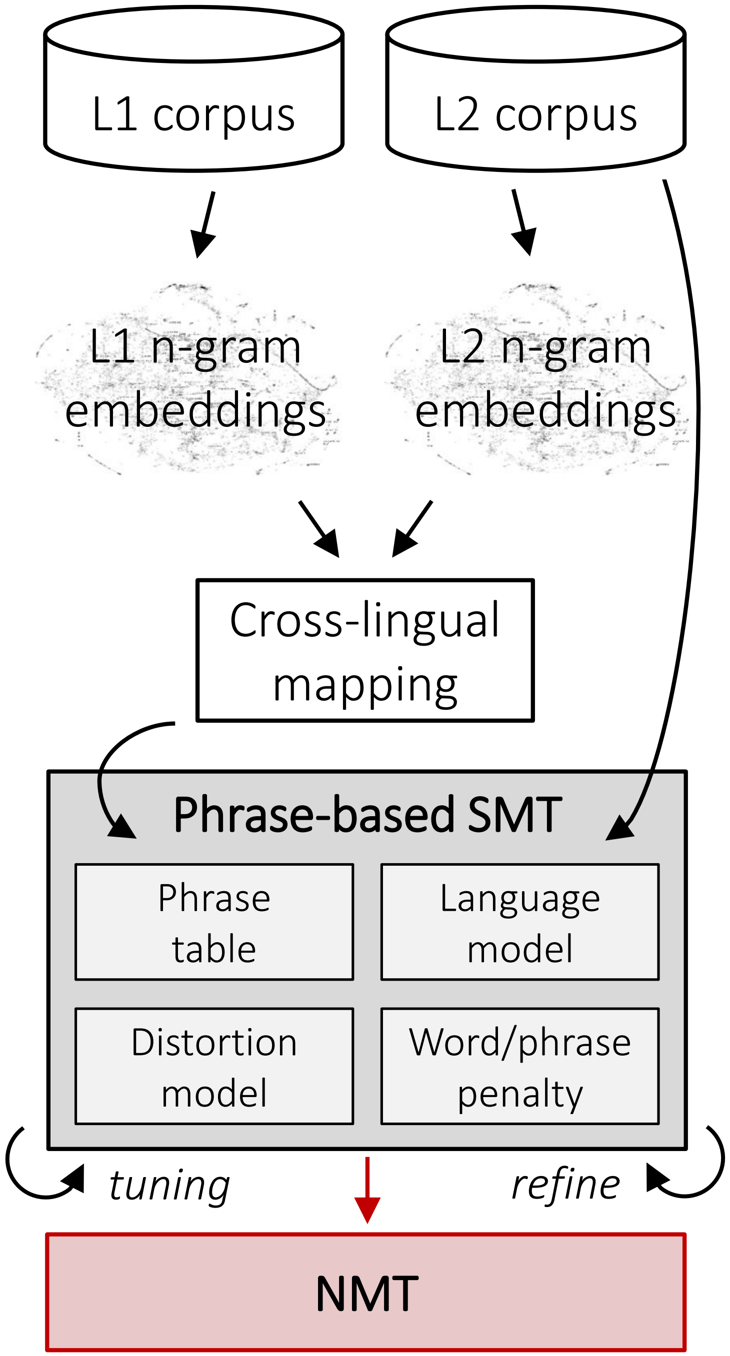


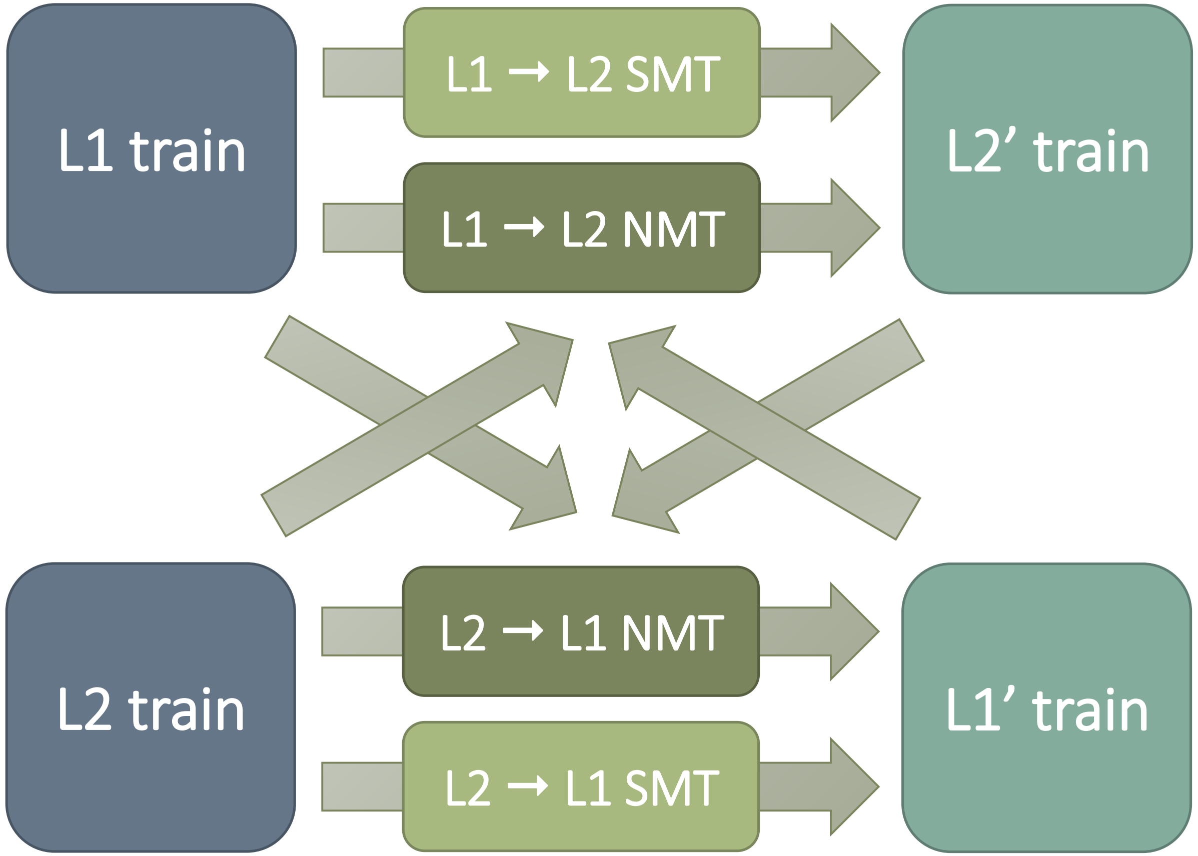
NMT hybridization



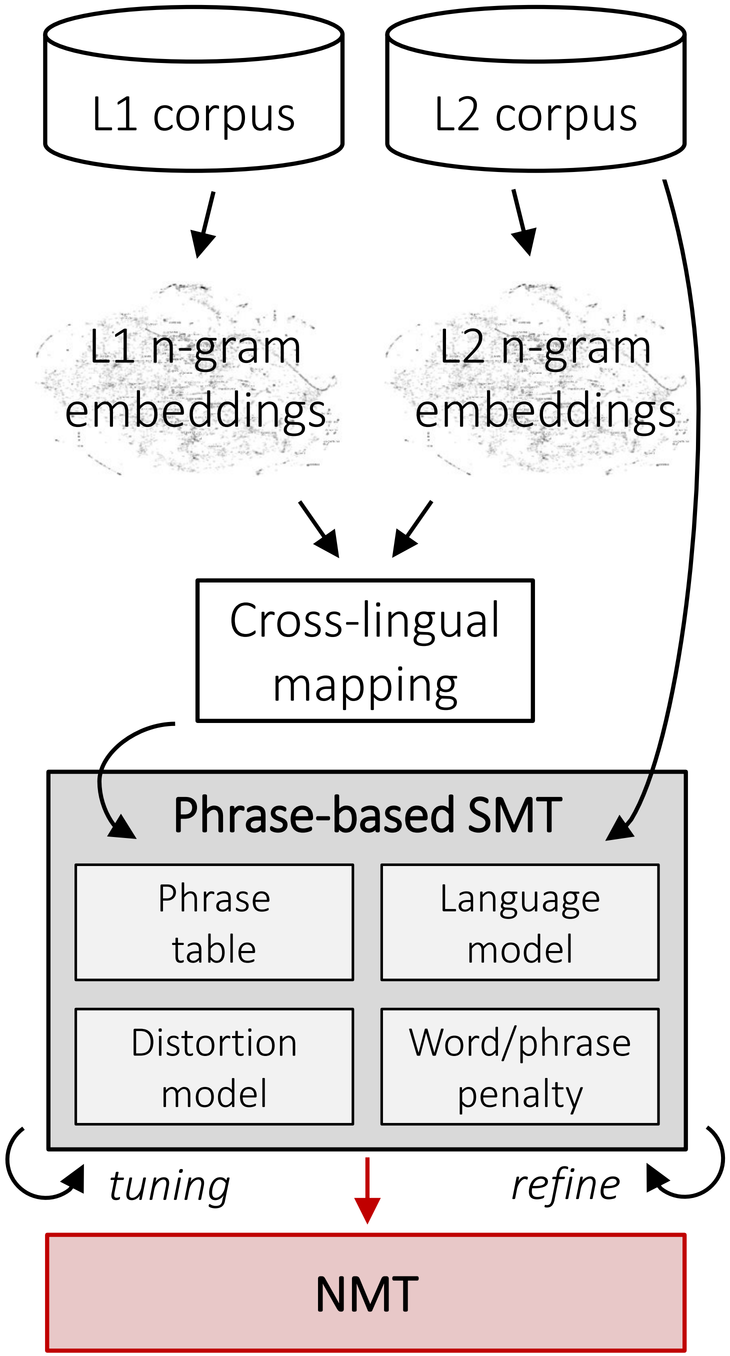


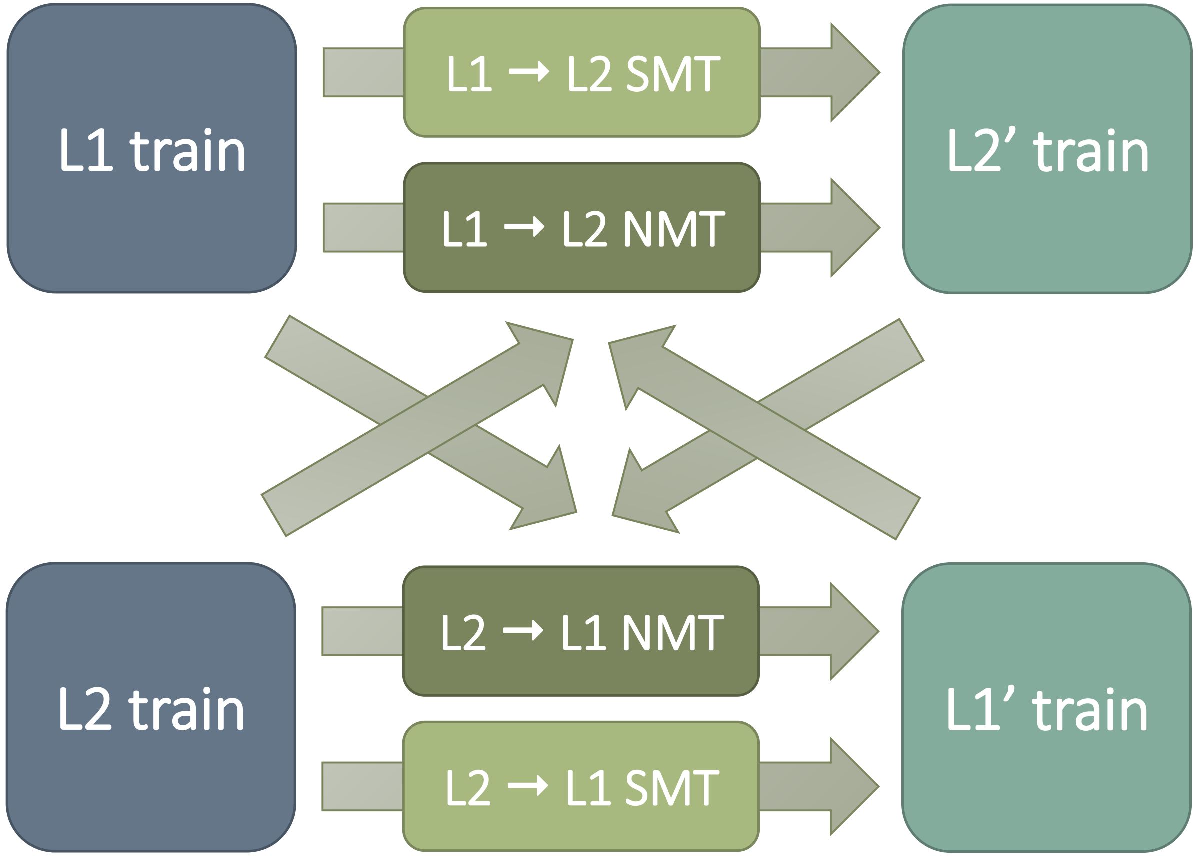
NMT hybridization





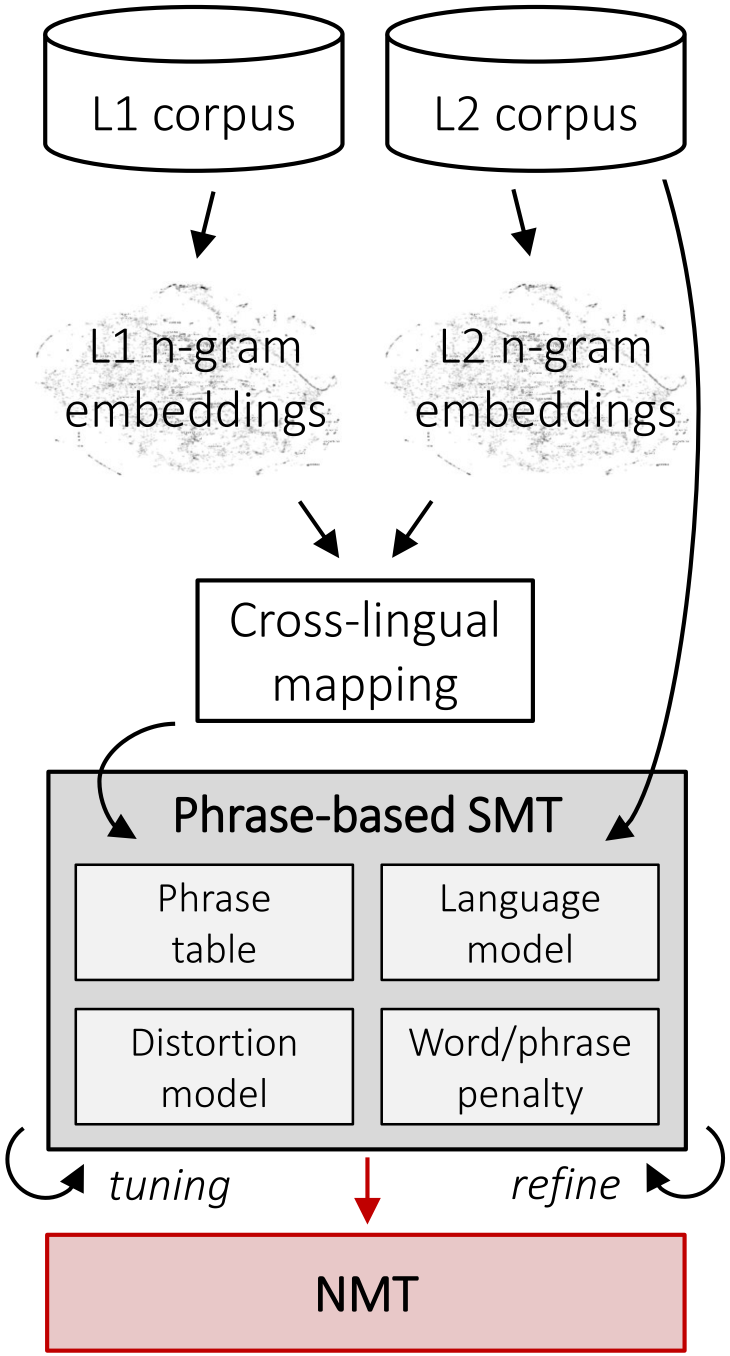
NMT hybridization

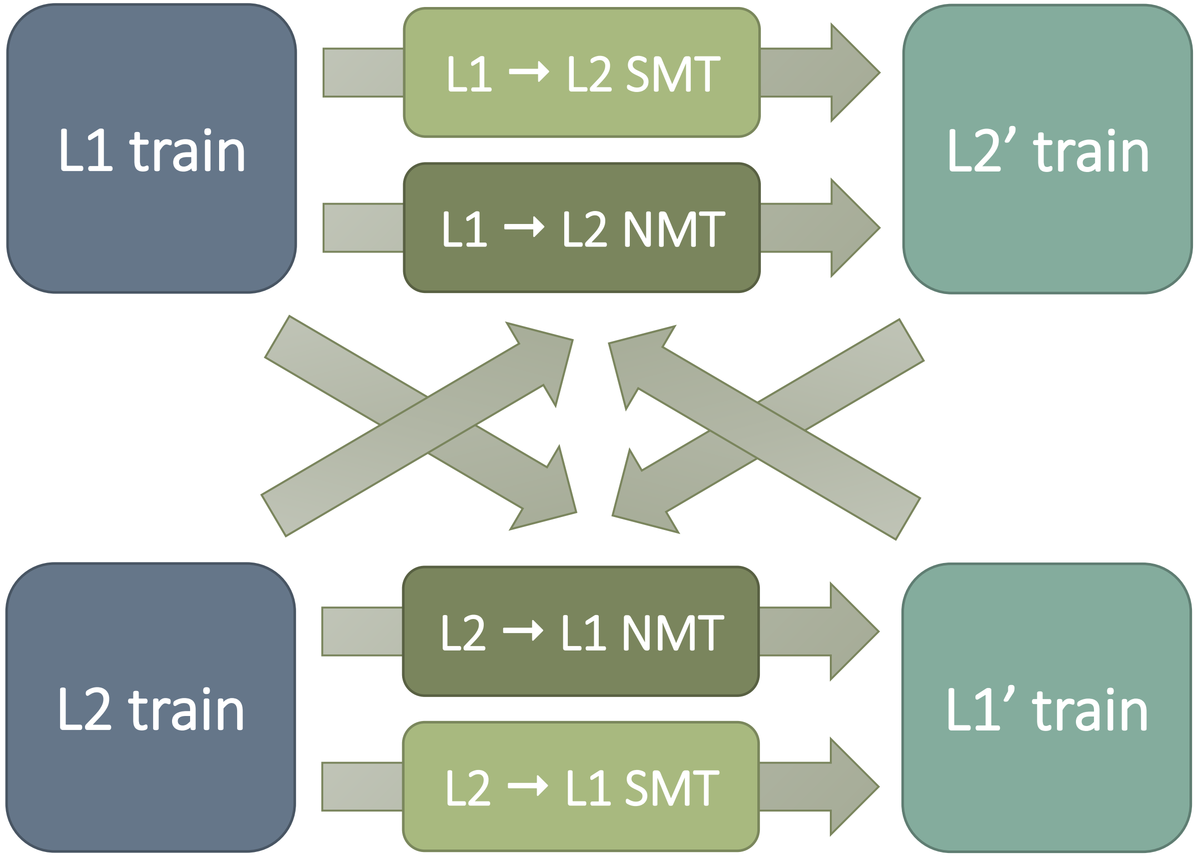




$$N_{SMT} = N \cdot \max(0, 1 - t/a)$$

NMT hybridization

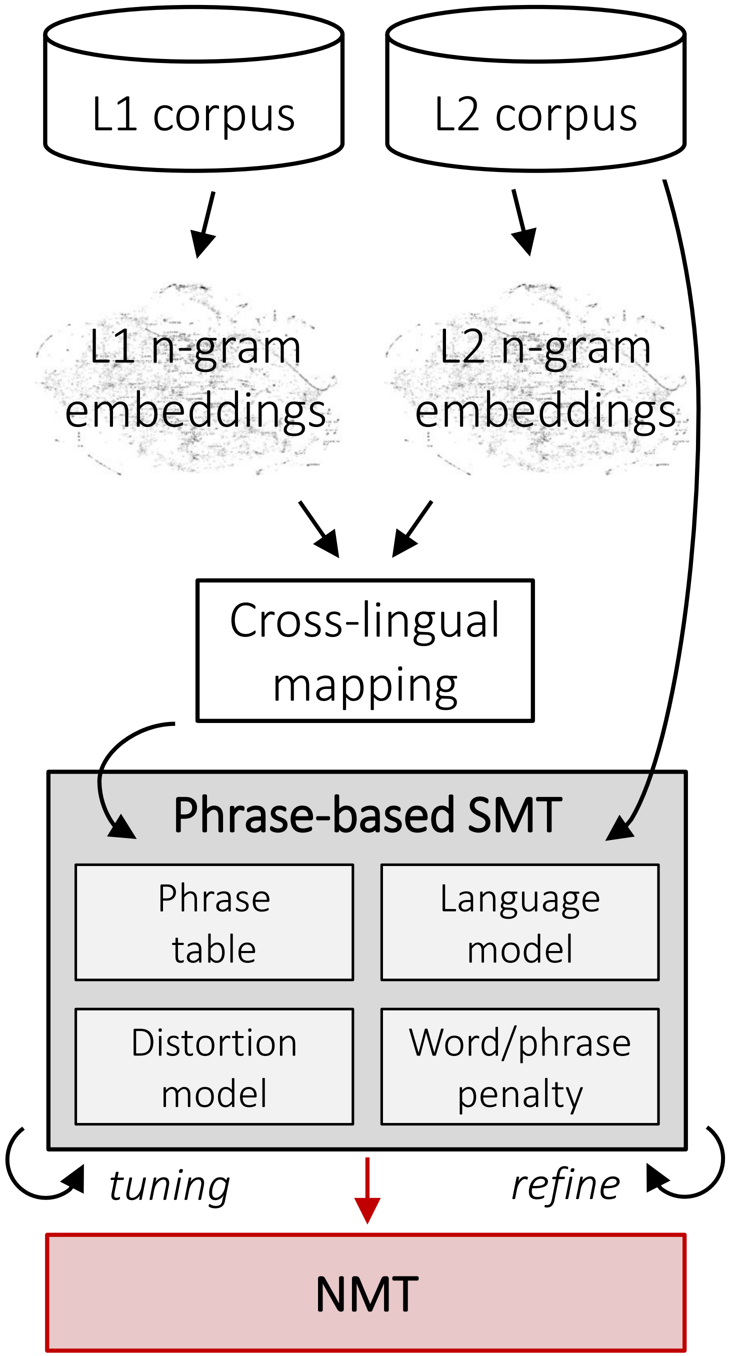




$$N_{SMT} = N \cdot \max(0, 1 - t/a)$$

$$N_{NMT} = N - N_{SMT}$$

NMT hybridization



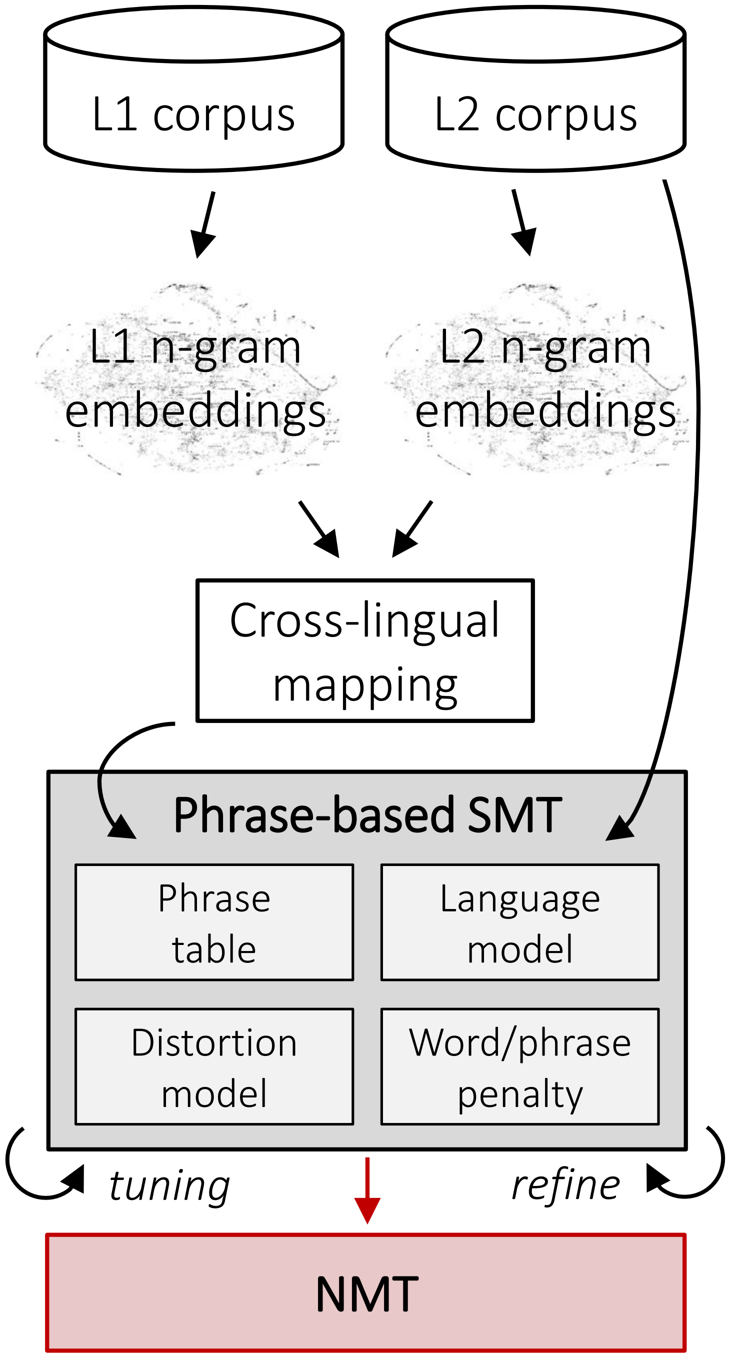
EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

	FR-EN	EN-FR	DE-EN	EN-DE
NMT (ICLR'18)*	15.6	15.1	10.2	6.6
Initial SMT (ACL'19)	22.4	19.6	15.3	11.0
+ Tuning	23.4	21.9	15.4	11.2
+ Refinement	27.9	27.8	19.7	14.7

*Tokenized BLEU (about 1-2 points higher)

NMT hybridization

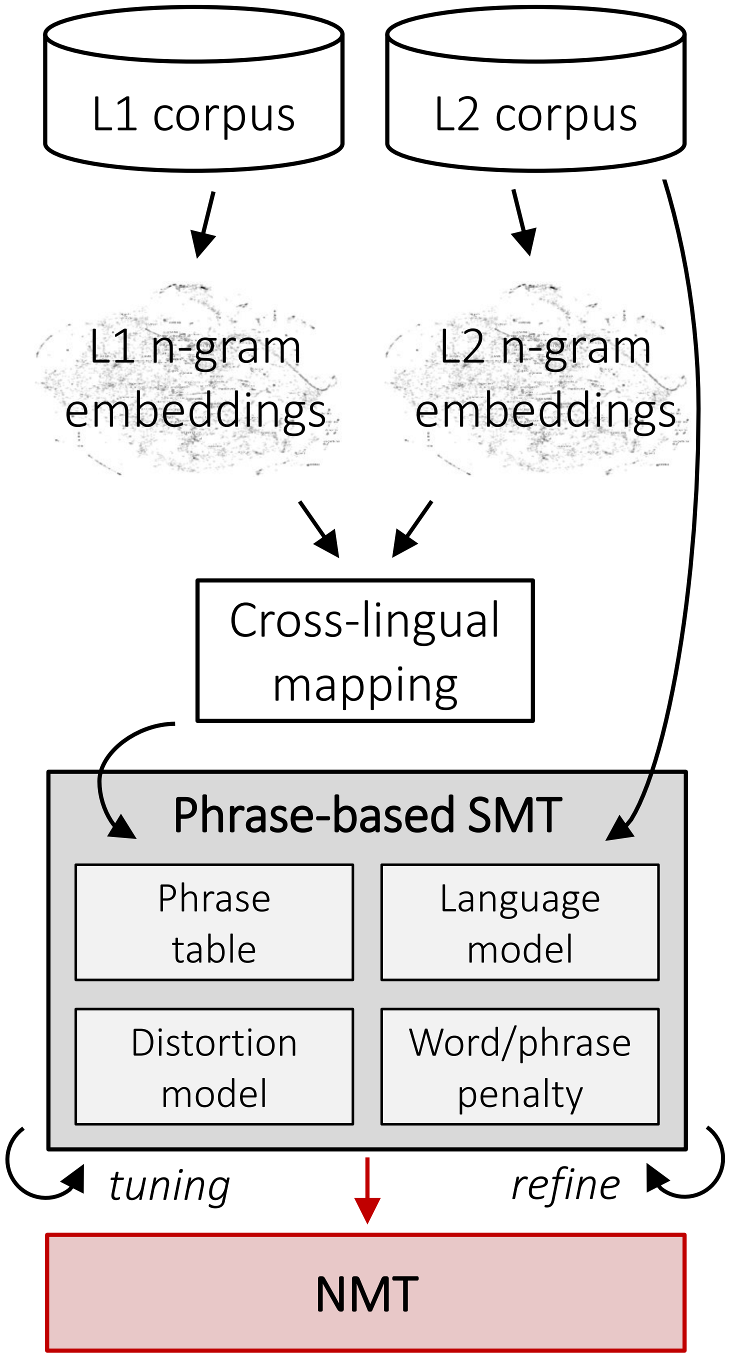


EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

	FR-EN	EN-FR	DE-EN	EN-DE
NMT (ICLR'18)*	15.6	15.1	10.2	6.6
Initial SMT (ACL'19)	22.4	19.6	15.3	11.0
+ Tuning	23.4	21.9	15.4	11.2
+ Refinement	27.9	27.8	19.7	14.7
+ NMT hybrid	33.2	33.6	26.4	21.2

*Tokenized BLEU (about 1-2 points higher)



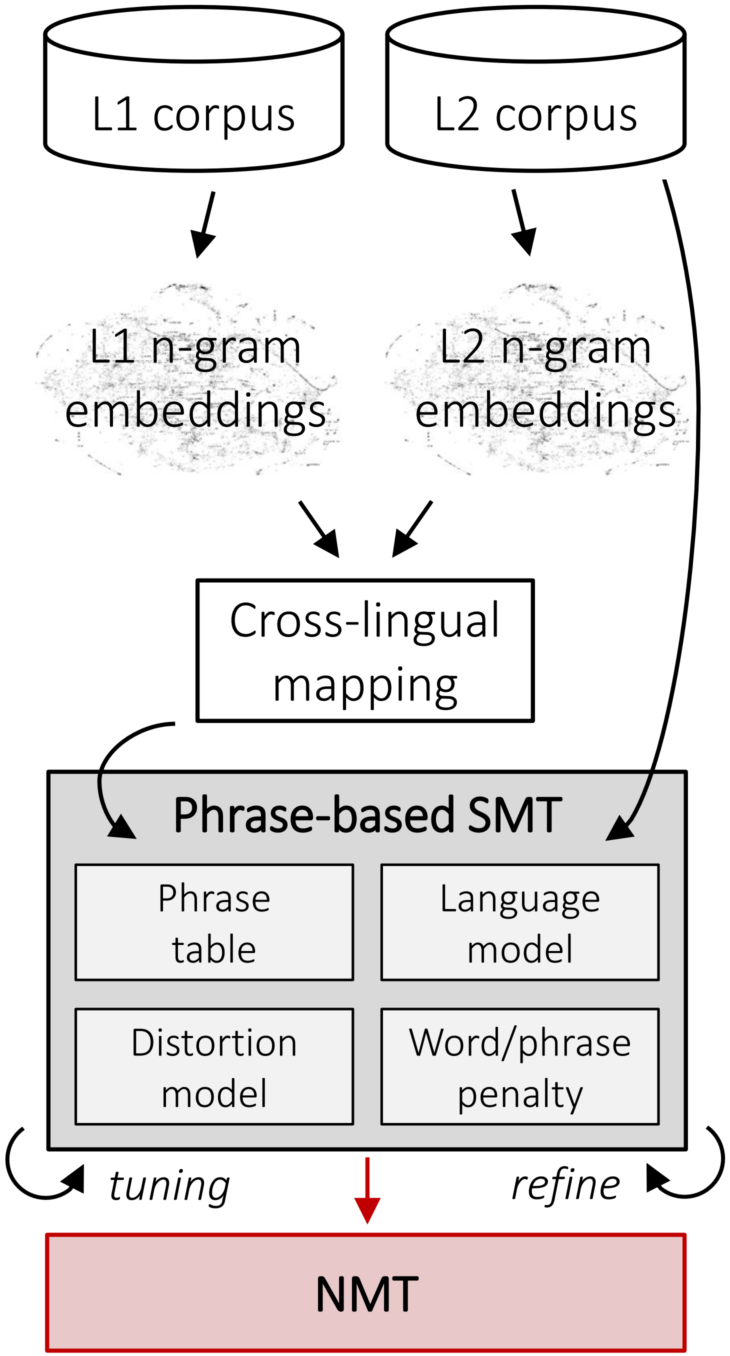
NMT hybridization

EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

	FR-EN	EN-FR	DE-EN	EN-DE
NMT (ICLR'18)*	15.6	15.1	10.2	6.6
Initial SMT (ACL'19)	22.4	19.6	15.3	11.0
+ Tuning	23.4	21.9	15.4	11.2
+ Refinement	27.9	27.8	19.7	14.7
+ NMT hybrid	33.2	33.6	26.4	21.2

*Tokenized BLEU (about 1-2 points higher)



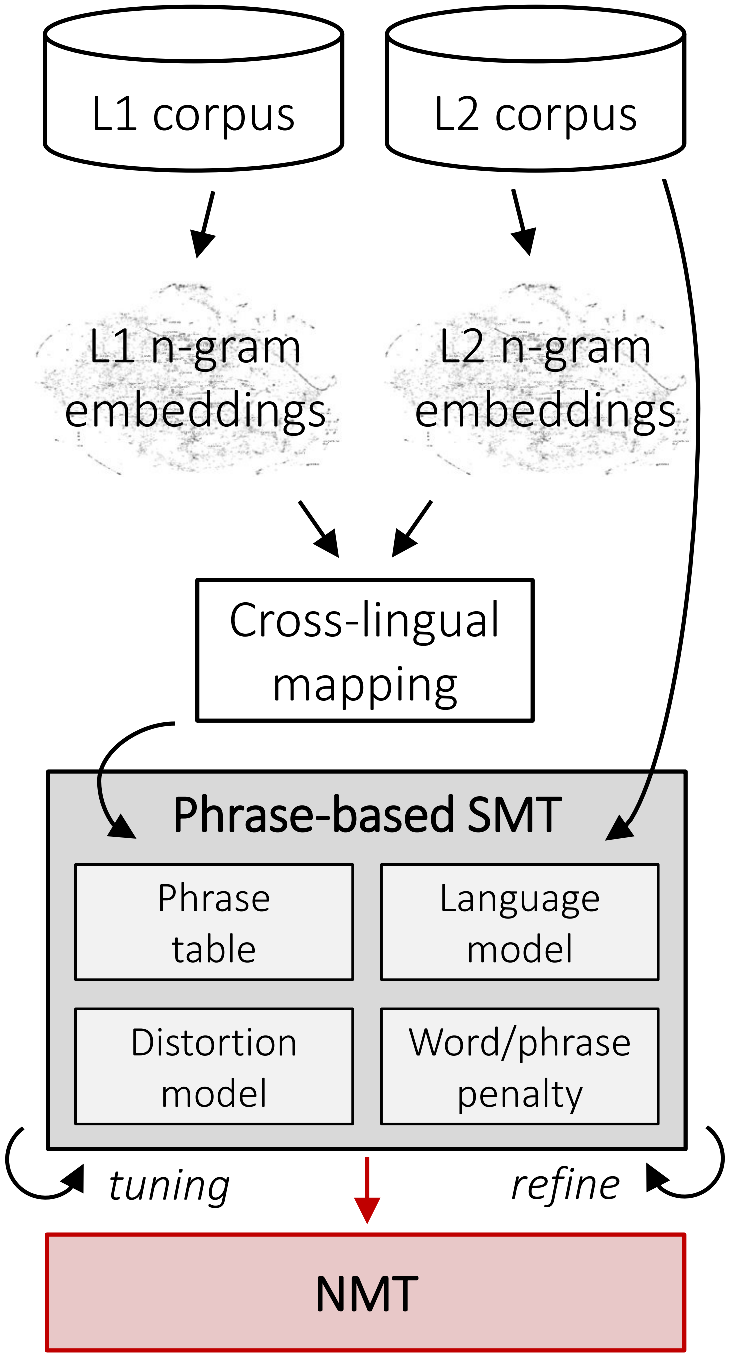
NMT hybridization

EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

		FR-EN	EN-FR	DE-EN	EN-DE
Unsup.	NMT (ICLR'18)*	15.6	15.1	10.2	6.6
	Initial SMT (ACL'19)	22.4	19.6	15.3	11.0
	+ Tuning	23.4	21.9	15.4	11.2
	+ Refinement	27.9	27.8	19.7	14.7
	+ NMT hybrid	33.2	33.6	26.4	21.2
Supervised					

*Tokenized BLEU (about 1-2 points higher)



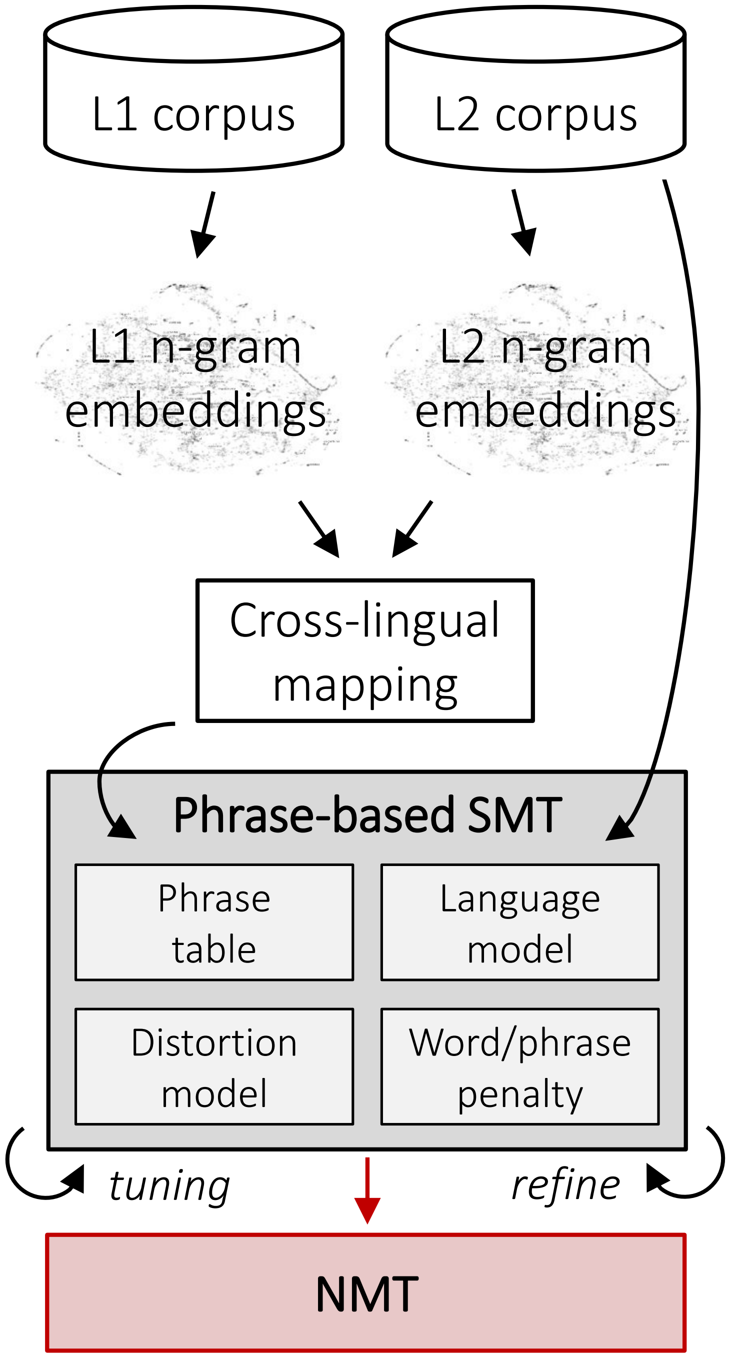
NMT hybridization

EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

		FR-EN	EN-FR	DE-EN	EN-DE
Unsup.	NMT (ICLR'18)*	15.6	15.1	10.2	6.6
	Initial SMT (ACL'19)	22.4	19.6	15.3	11.0
	+ Tuning	23.4	21.9	15.4	11.2
	+ Refinement	27.9	27.8	19.7	14.7
	+ NMT hybrid	33.2	33.6	26.4	21.2
Supervised	WMT winner	35.0	35.8	29.0	20.6

*Tokenized BLEU (about 1-2 points higher)



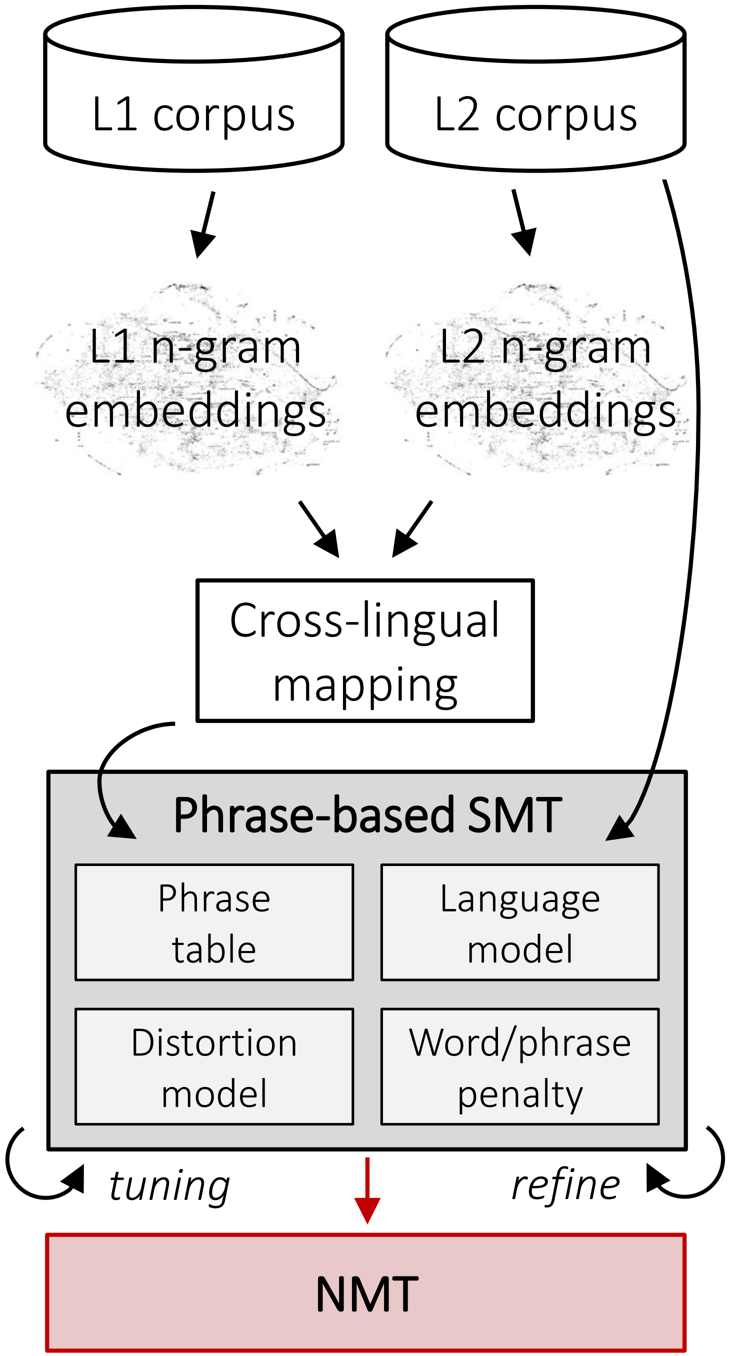
NMT hybridization

EXPERIMENTS

- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

		FR-EN	EN-FR	DE-EN	EN-DE
Unsup.	NMT (ICLR'18)*	15.6	15.1	10.2	6.6
	Initial SMT (ACL'19)	22.4	19.6	15.3	11.0
	+ Tuning	23.4	21.9	15.4	11.2
	+ Refinement	27.9	27.8	19.7	14.7
	+ NMT hybrid	33.2	33.6	26.4	21.2
Supervised	WMT winner	35.0	35.8	29.0	20.6
	Vaswani et al. (NIPS'17)*	-	41.0	-	28.4

*Tokenized BLEU (about 1-2 points higher)



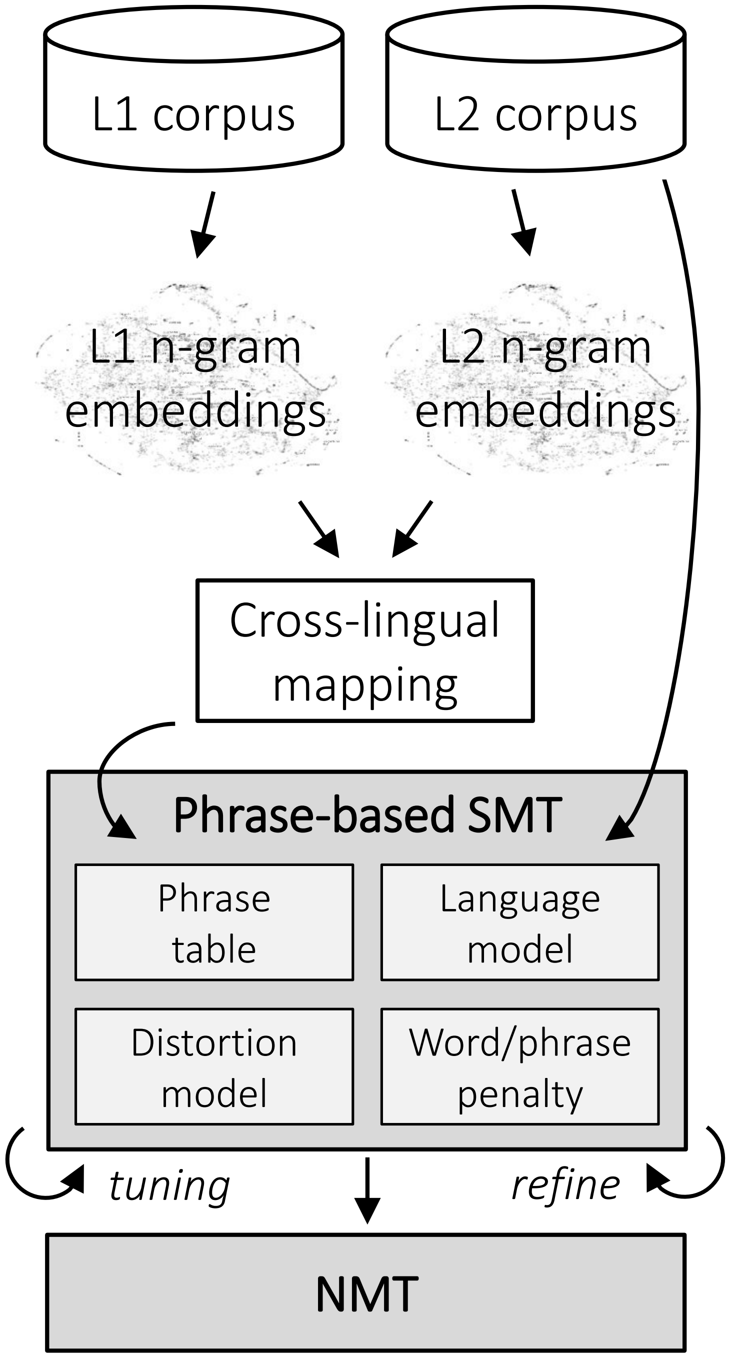
NMT hybridization

EXPERIMENTS

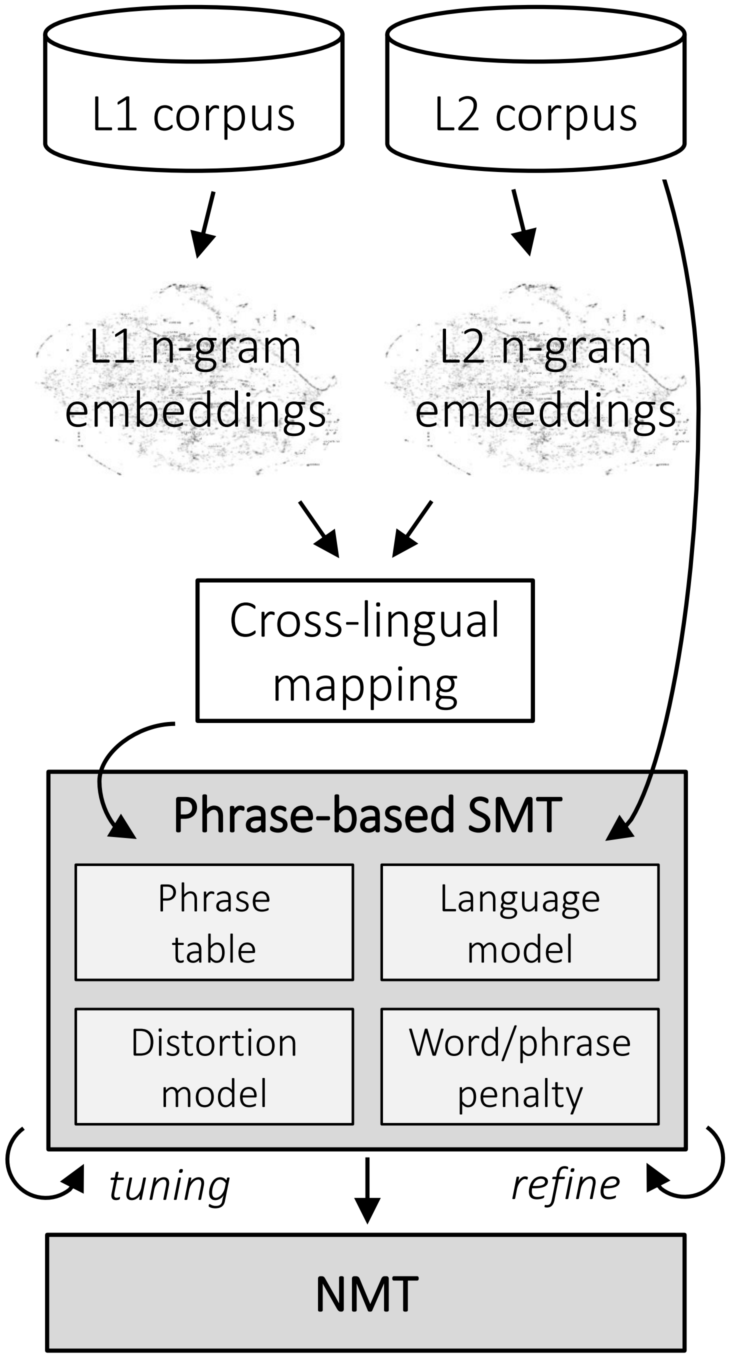
- Languages: French-English, German-English
- Training: WMT-14 News Crawl
- Test set: WMT-14 newstest (BLEU)

		FR-EN	EN-FR	DE-EN	EN-DE
Unsup.	NMT (ICLR'18)*	15.6	15.1	10.2	6.6
	Initial SMT (ACL'19)	22.4	19.6	15.3	11.0
	+ Tuning	23.4	21.9	15.4	11.2
	+ Refinement	27.9	27.8	19.7	14.7
	+ NMT hybrid	33.2	33.6	26.4	21.2
Supervised	WMT winner	35.0	35.8	29.0	20.6
	Vaswani et al. (NIPS'17)*	-	41.0	-	28.4
	Edunov et al. (EMNLP'18)	-	43.8	-	33.8

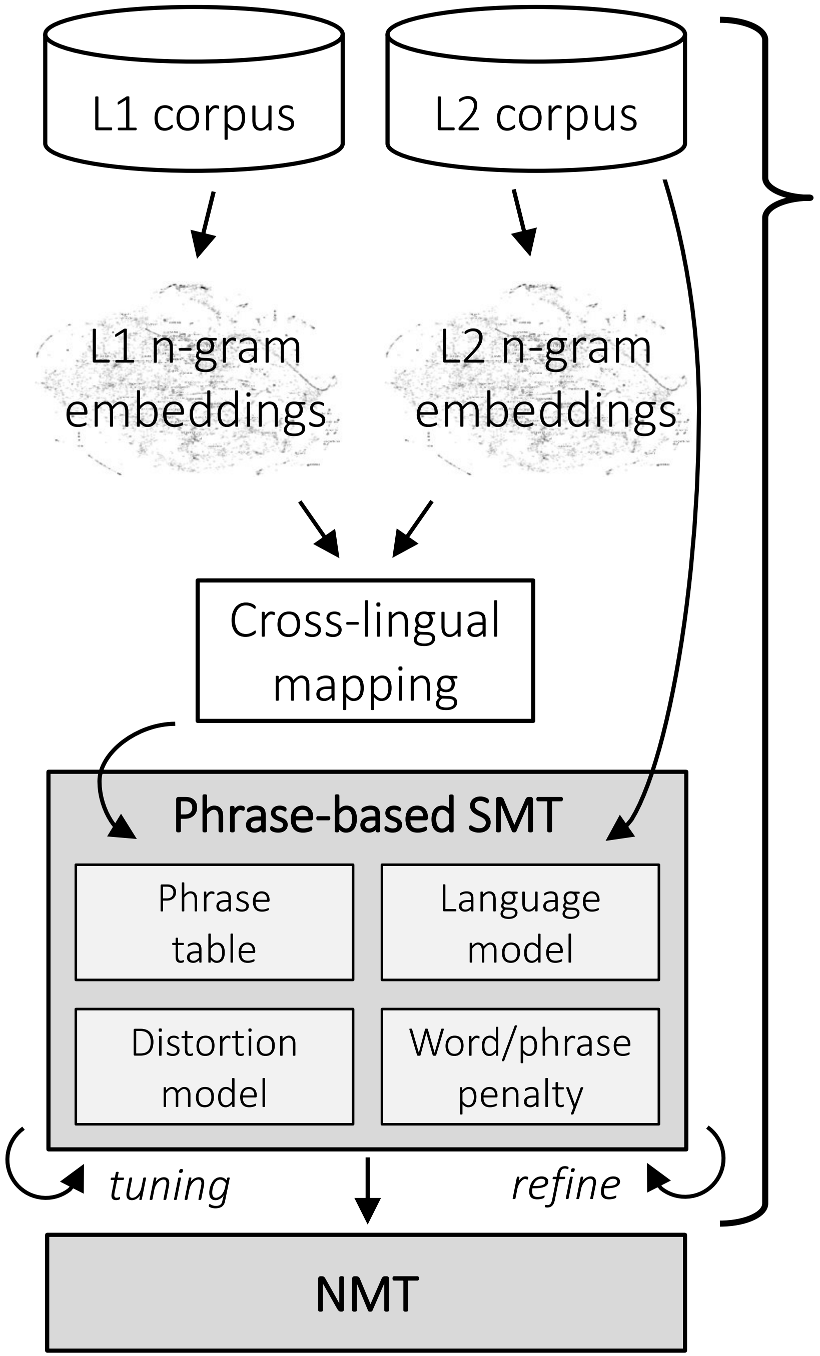
*Tokenized BLEU (about 1-2 points higher)


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine
NMT

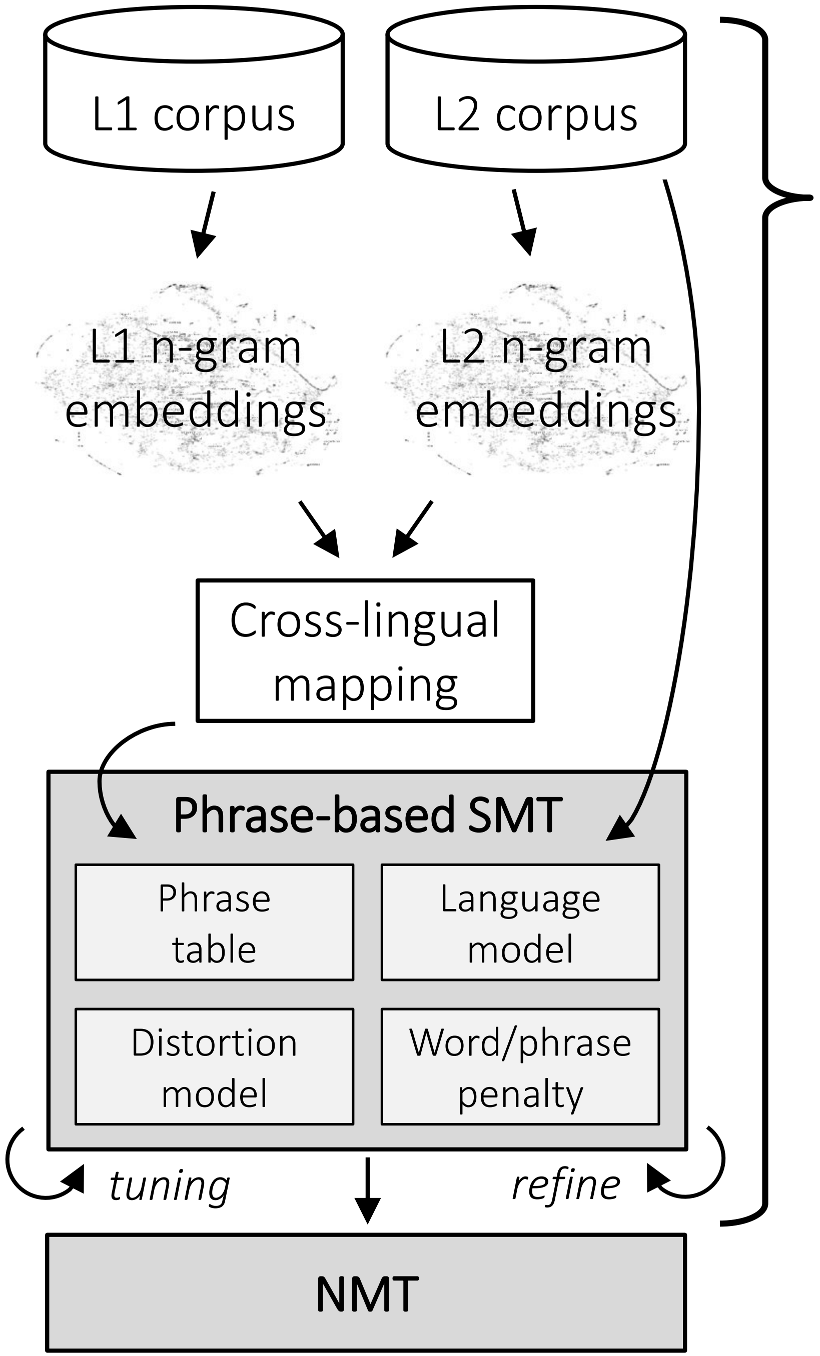
What's next?


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine
NMT

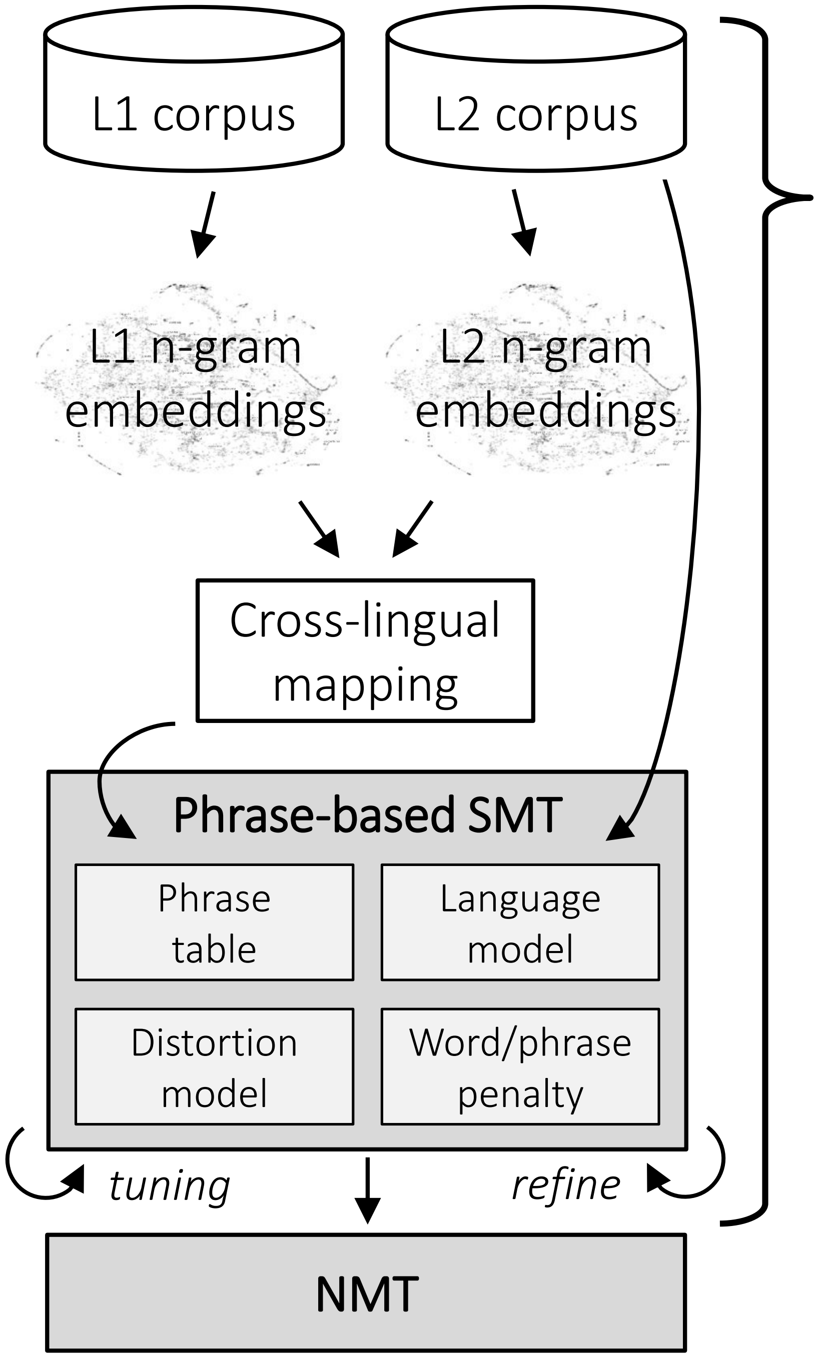
What's next?


L1 corpus
L2 corpus
L1 n-gram embeddings
L2 n-gram embeddings
Cross-lingual mapping
Phrase-based SMT
Phrase table
Language model
Distortion model
Word/phrase penalty
tuning
refine
NMT

What's next?



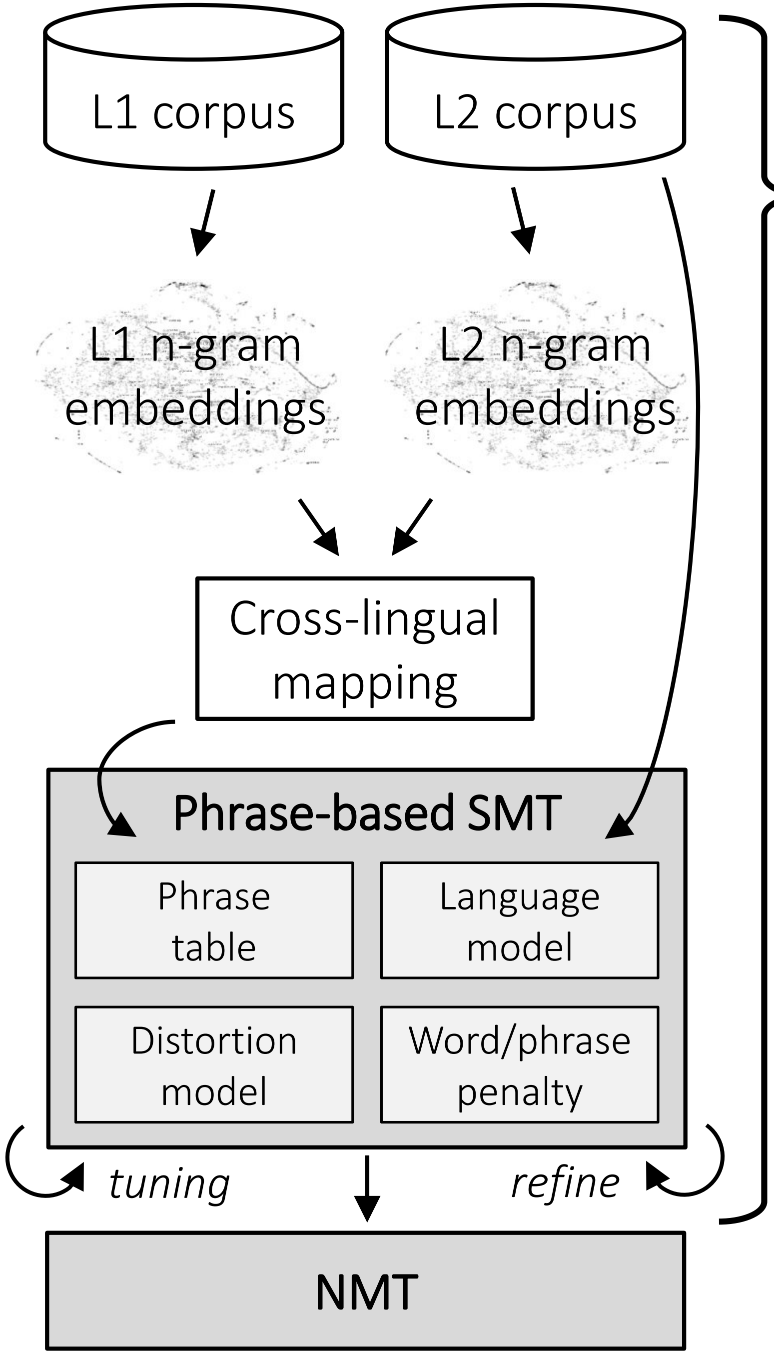
Do we really need this complex SMT part?



What's next?

Do we really need this complex SMT part?

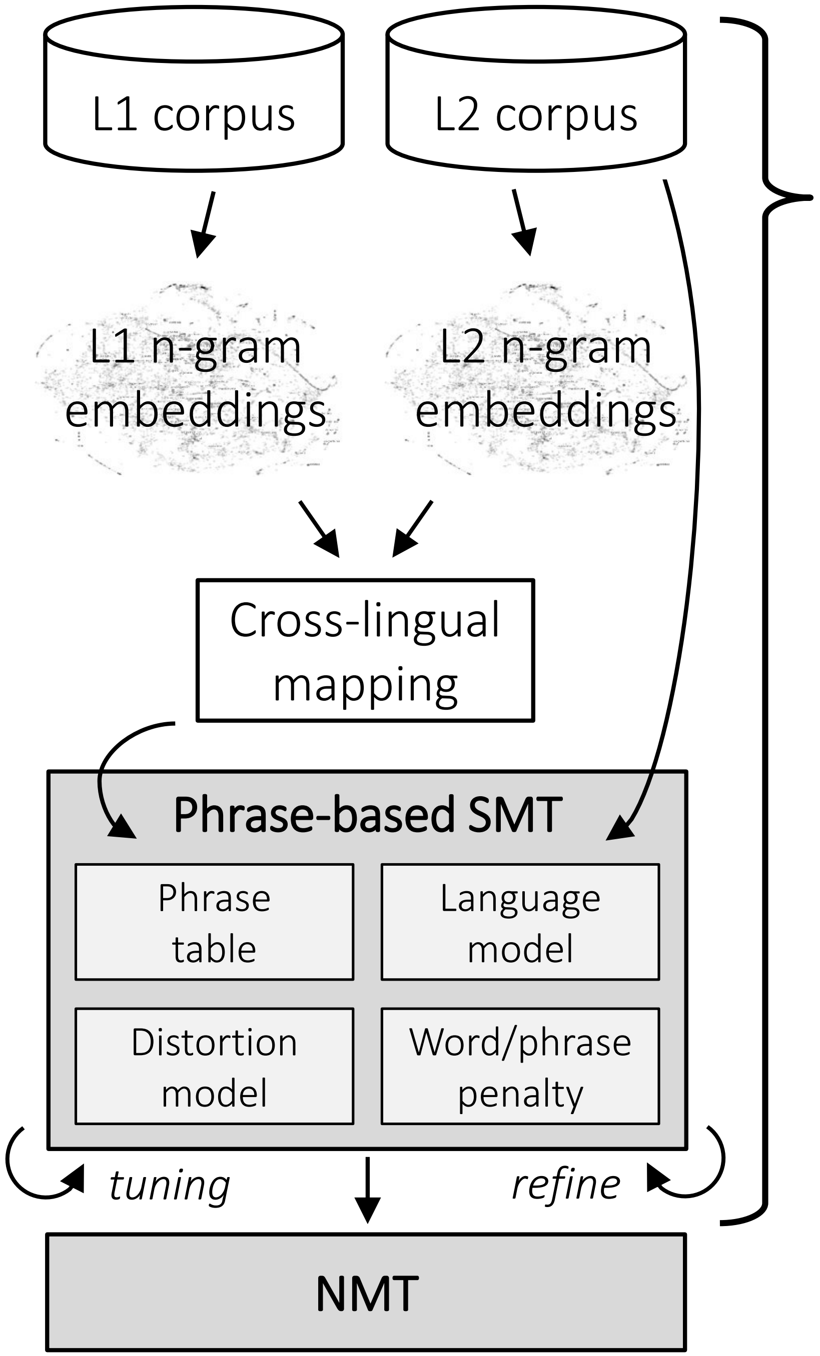
- We are (more or less) at par with supervised SMT, so how much can SMT help to keep making progress?



What's next?

Do we really need this complex SMT part?

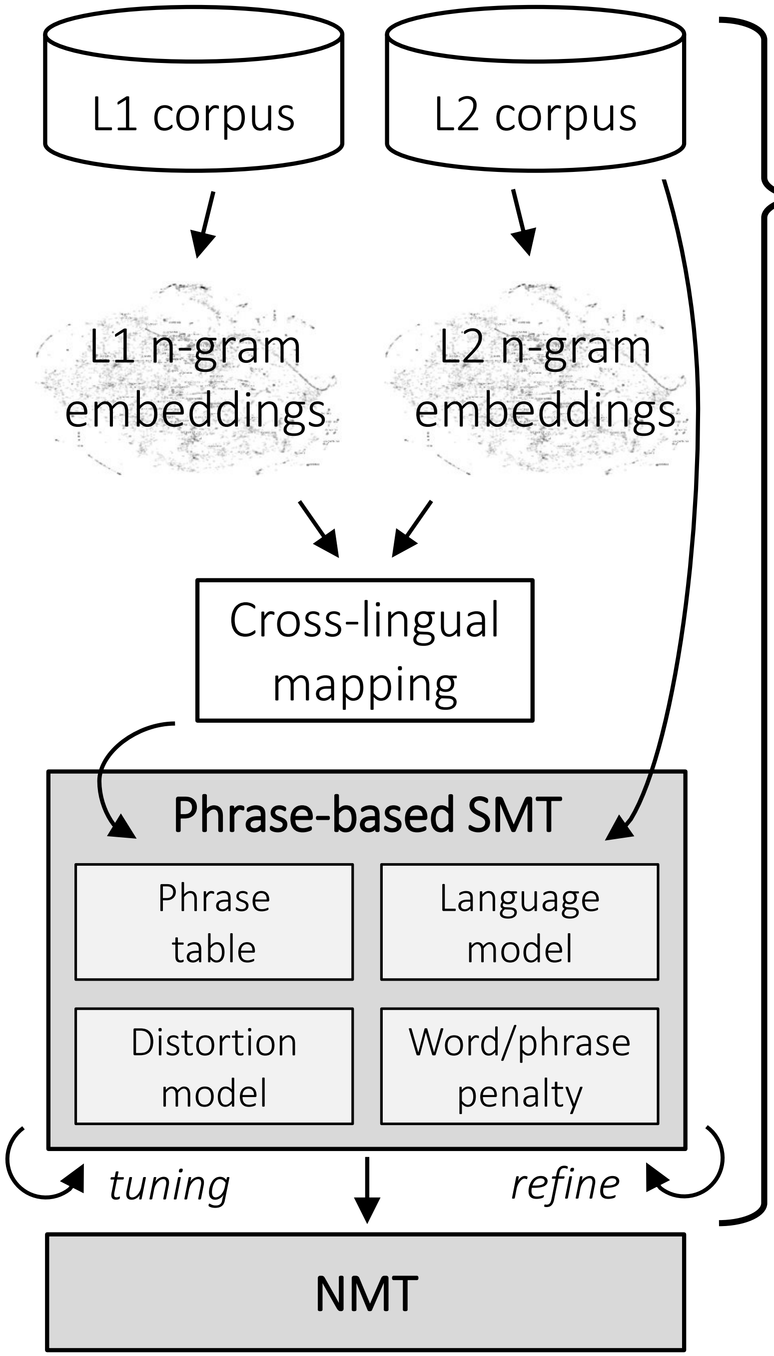
- We are (more or less) at par with supervised SMT, so how much can SMT help to keep making progress?



What's next?

Do we really need this complex SMT part?

- We are (more or less) at par with supervised SMT, so how much can SMT help to keep making progress?



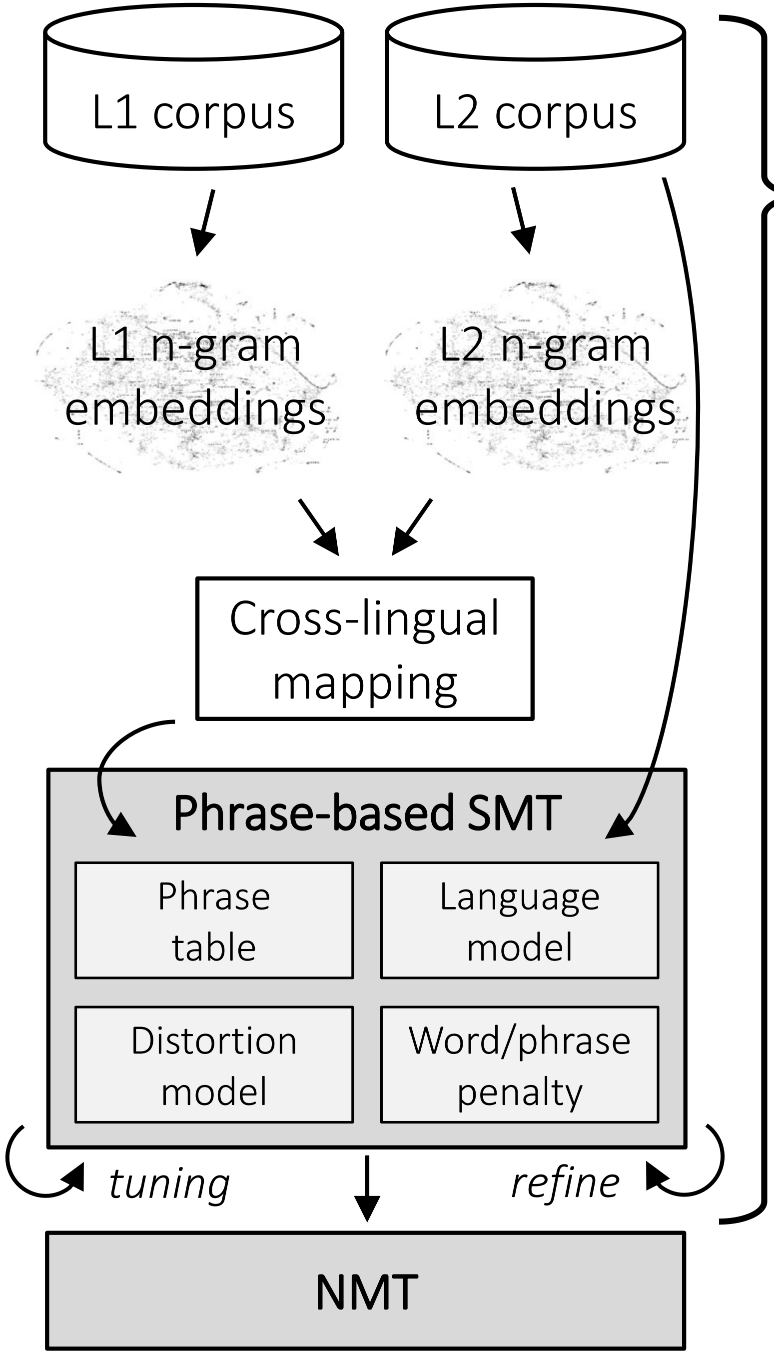
What's next?

Do we really need this complex SMT part?

- We are (more or less) at par with supervised SMT, so how much can SMT help to keep making progress?

I HAVE SOMETHING TO SAY ON THIS!

Lample & Conneau (arXiv'19)



What's next?

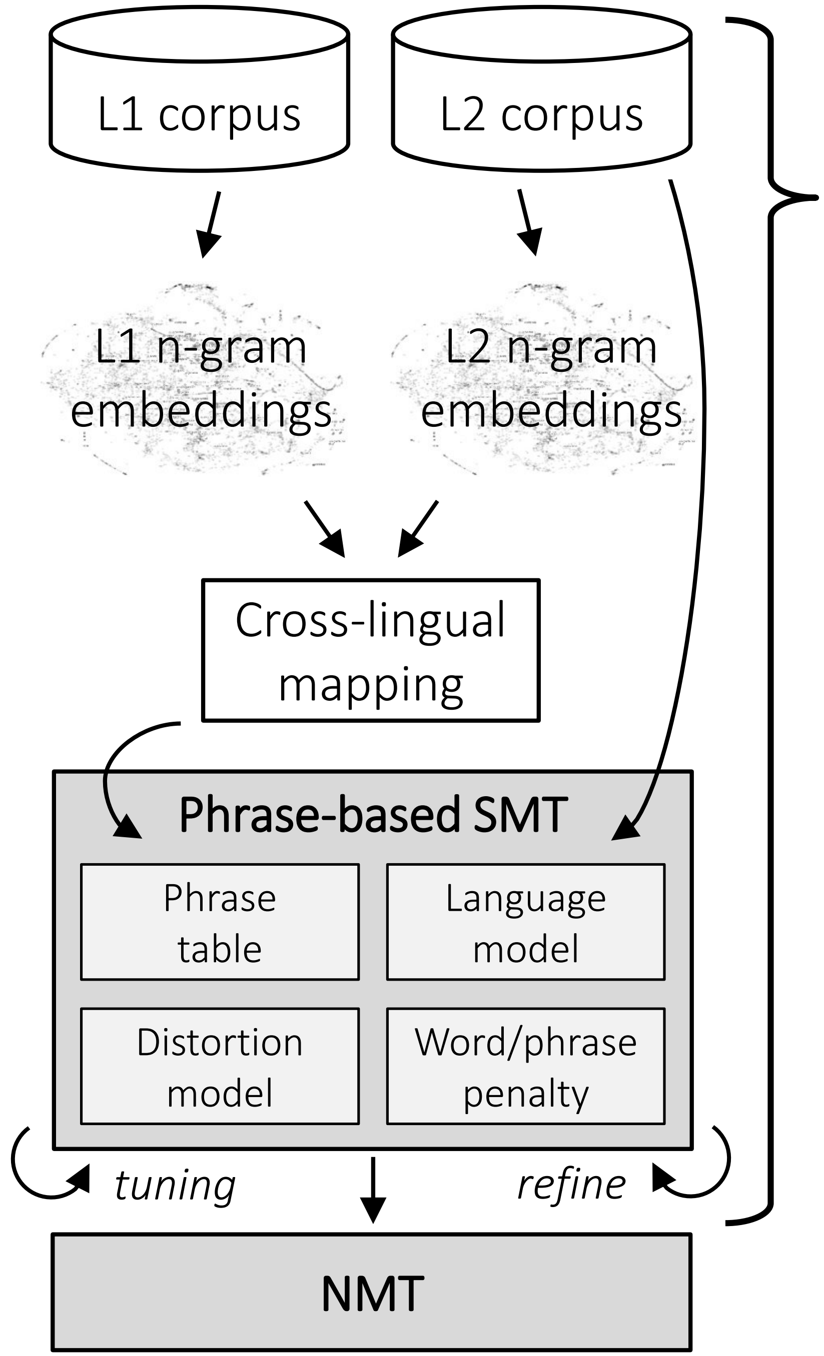
Do we really need this complex SMT part?

- We are (more or less) at par with supervised SMT, so how much can SMT help to keep making progress?

I HAVE SOMETHING TO SAY ON THIS!

Lampple & Conneau (arXiv'19)

- Multilingual BERT for initialization (joint masked language modeling over concatenated monolingual corpora)



What's next?

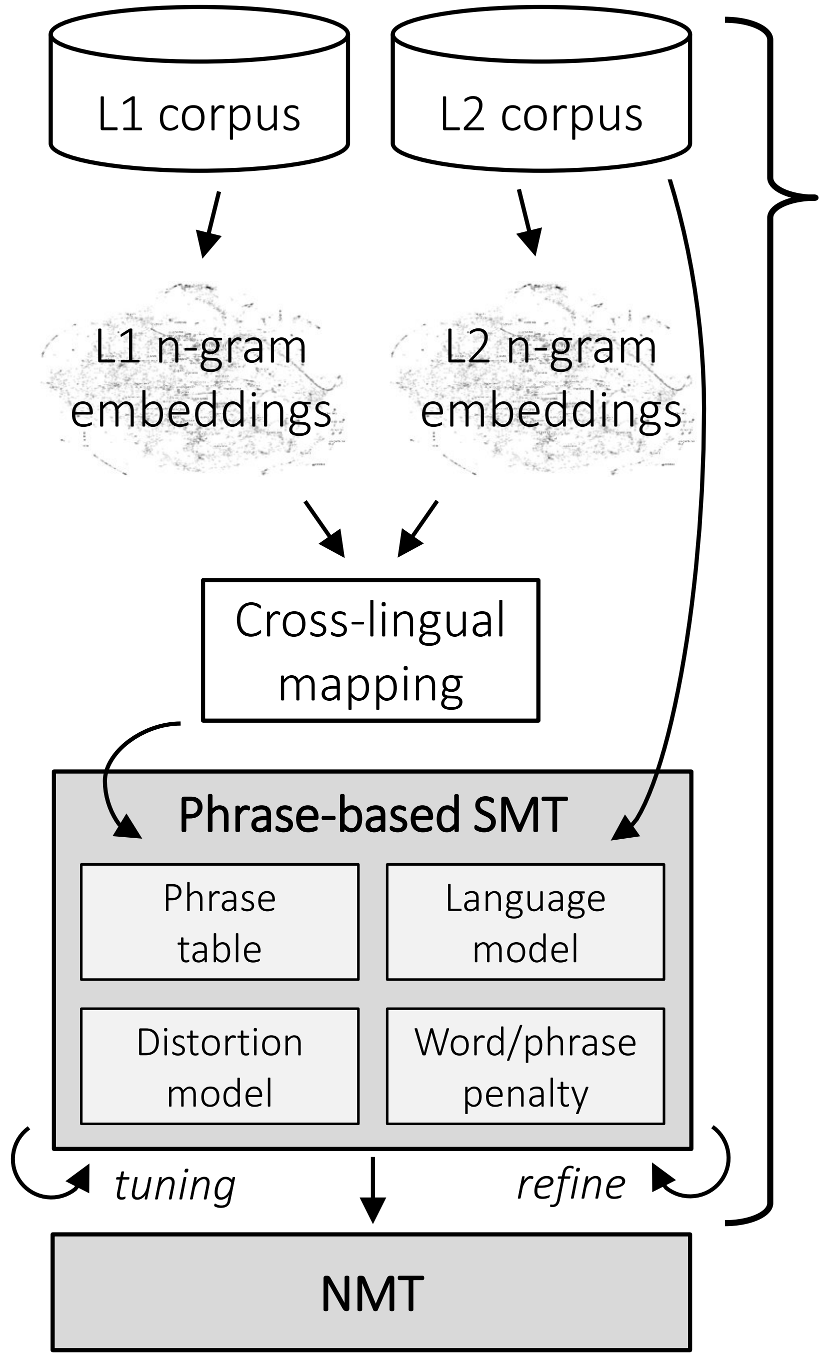
Do we really need this complex SMT part?

- We are (more or less) at par with supervised SMT, so how much can SMT help to keep making progress?

I HAVE SOMETHING TO SAY ON THIS!

Lample & Conneau (arXiv'19)

- Multilingual BERT for initialization (joint masked language modeling over concatenated monolingual corpora)
- Similar results to our final system



What's next?

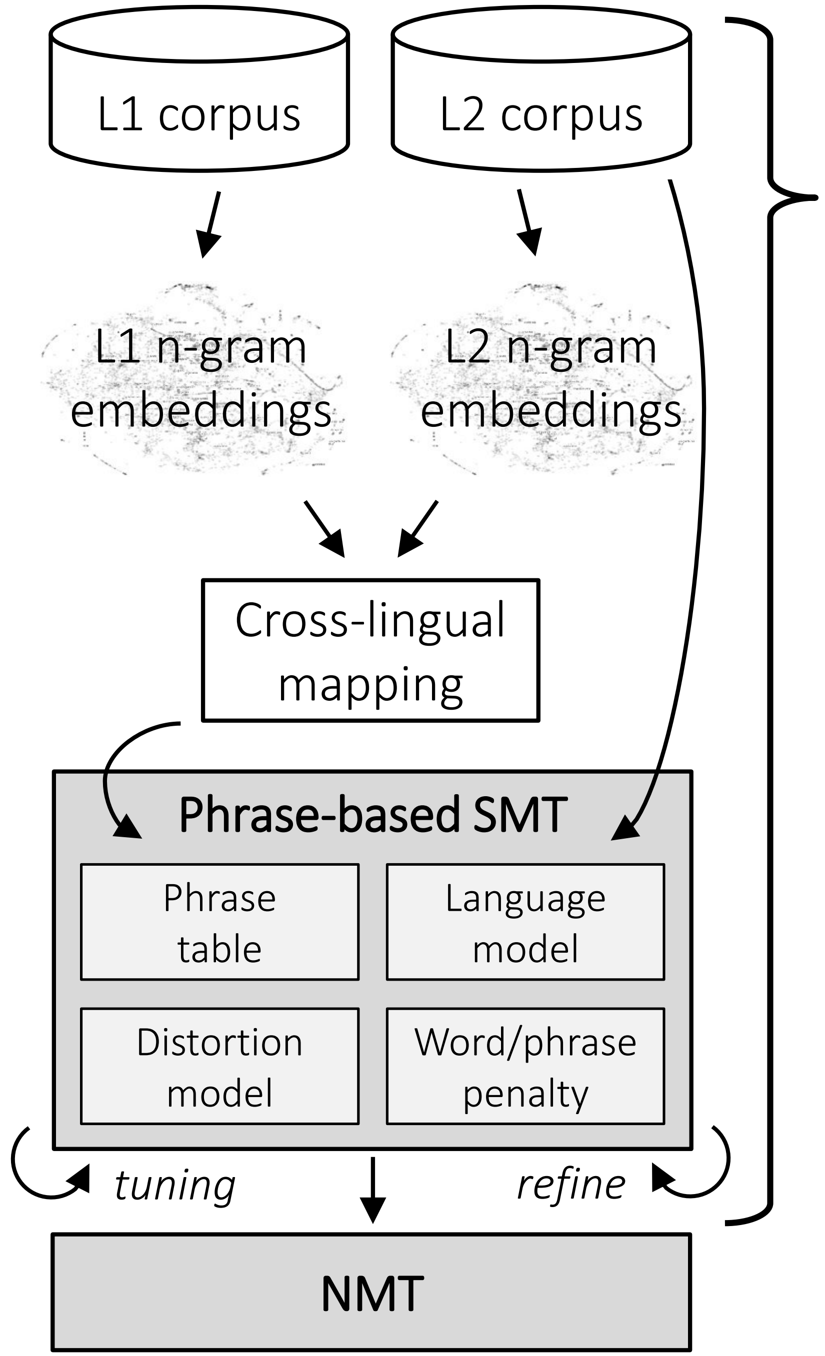
Do we really need this complex SMT part?

- We are (more or less) at par with supervised SMT, so how much can SMT help to keep making progress?

I HAVE SOMETHING TO SAY ON THIS!

Lample & Conneau (arXiv'19)

- Multilingual BERT for initialization (joint masked language modeling over concatenated monolingual corpora)
- Similar results to our final system
- Conceptually simpler, but computationally expensive



What's next?

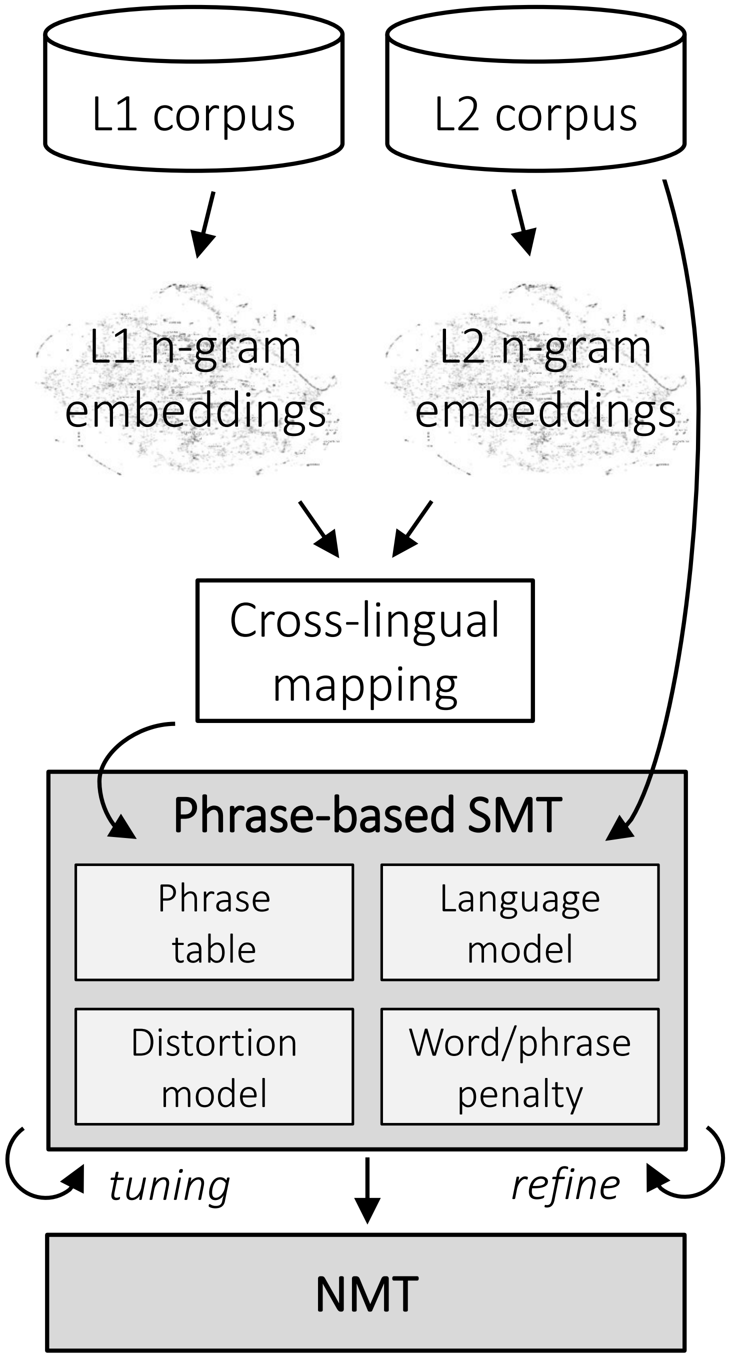
Do we really need this complex SMT part?

- We are (more or less) at par with supervised SMT, so how much can SMT help to keep making progress?

I HAVE SOMETHING TO SAY ON THIS!

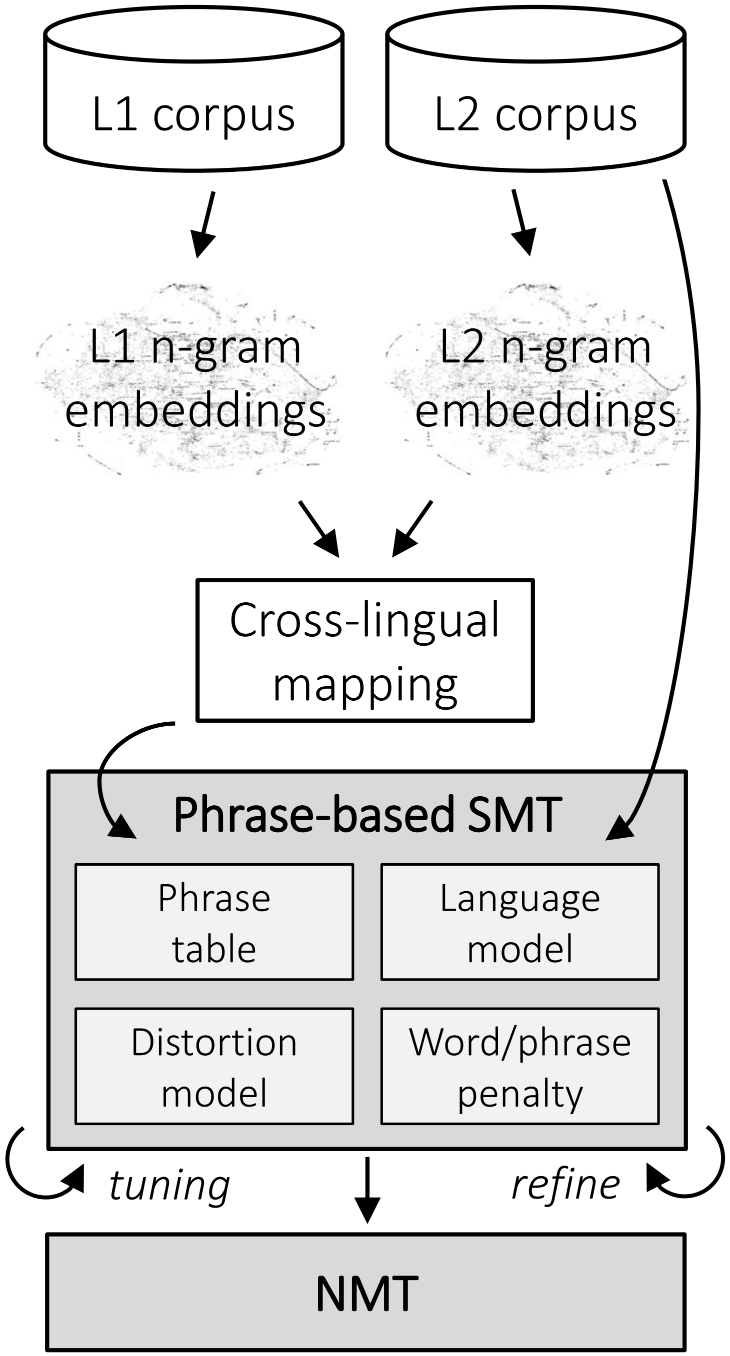
Lampple & Conneau (arXiv'19)

- Multilingual BERT for initialization (joint masked language modeling over concatenated monolingual corpora)
- Similar results to our final system
- Conceptually simpler, but computationally expensive
- Both approaches might be complementary!



What's next?

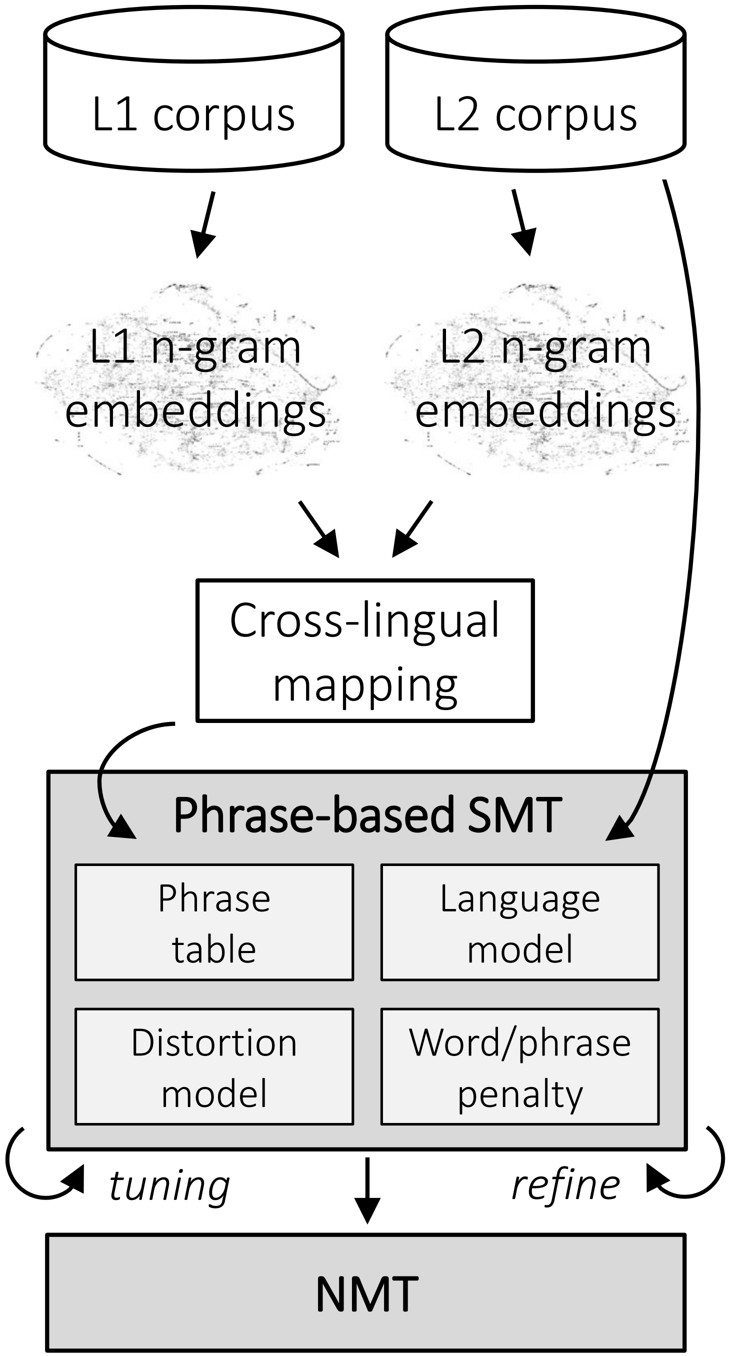
Some thoughts



What's next?

Some thoughts

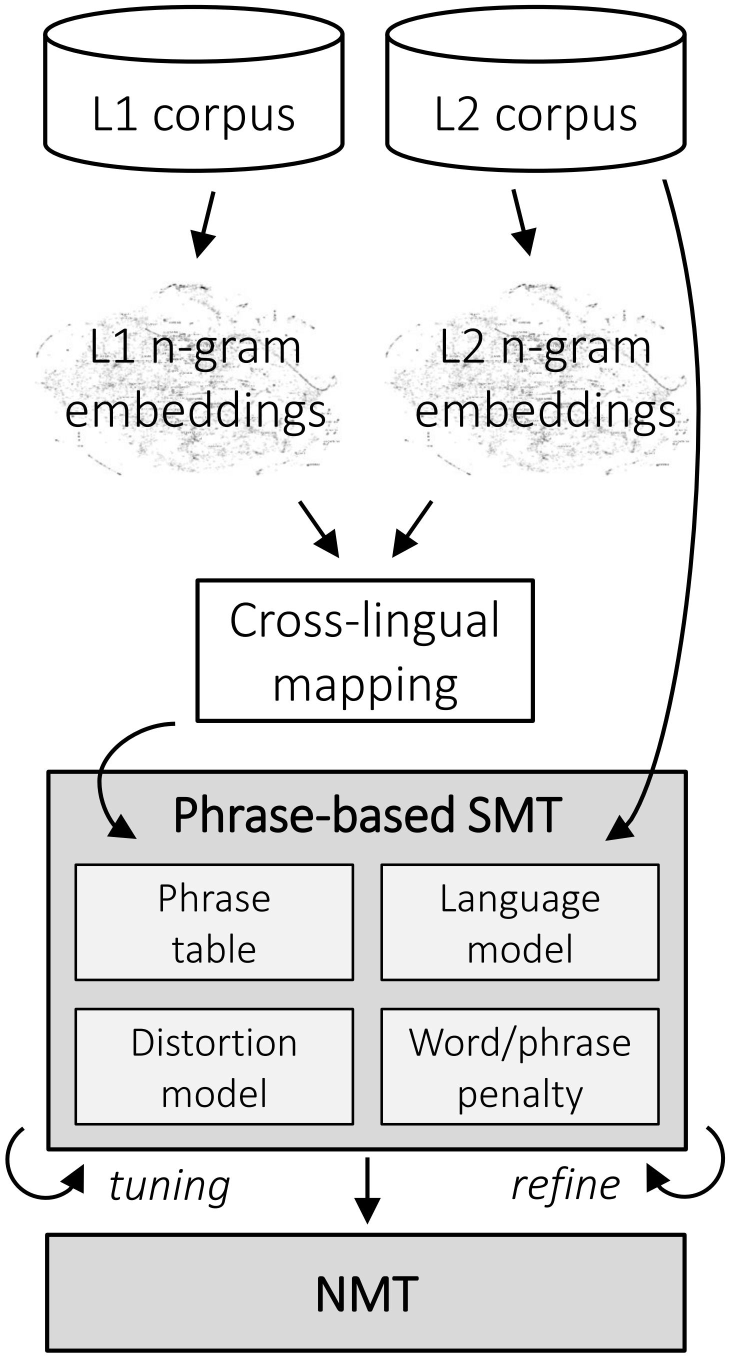
- Main recipe: iterative back-translation + smart initialization



What's next?

Some thoughts

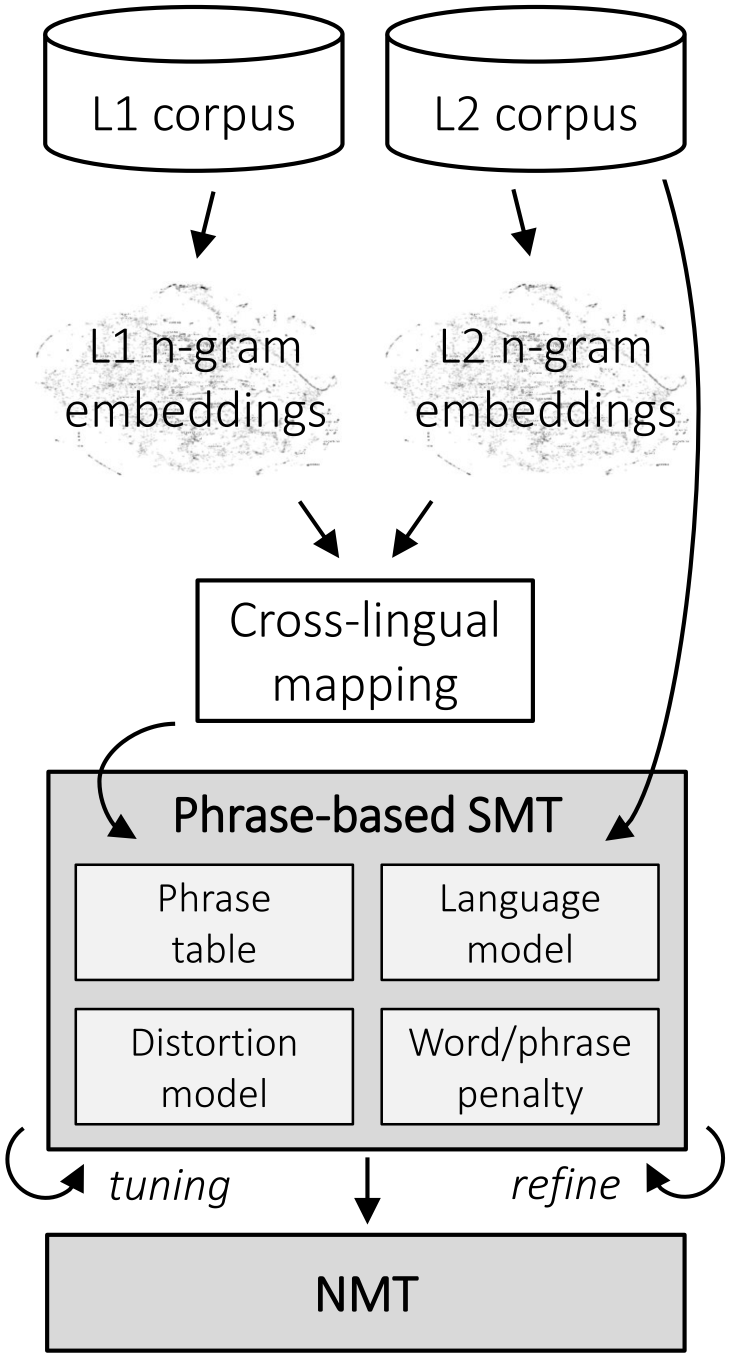
- Main recipe: iterative back-translation + smart initialization
  - Further explore the specifics? Start thinking beyond this paradigm?



What's next?

Some thoughts

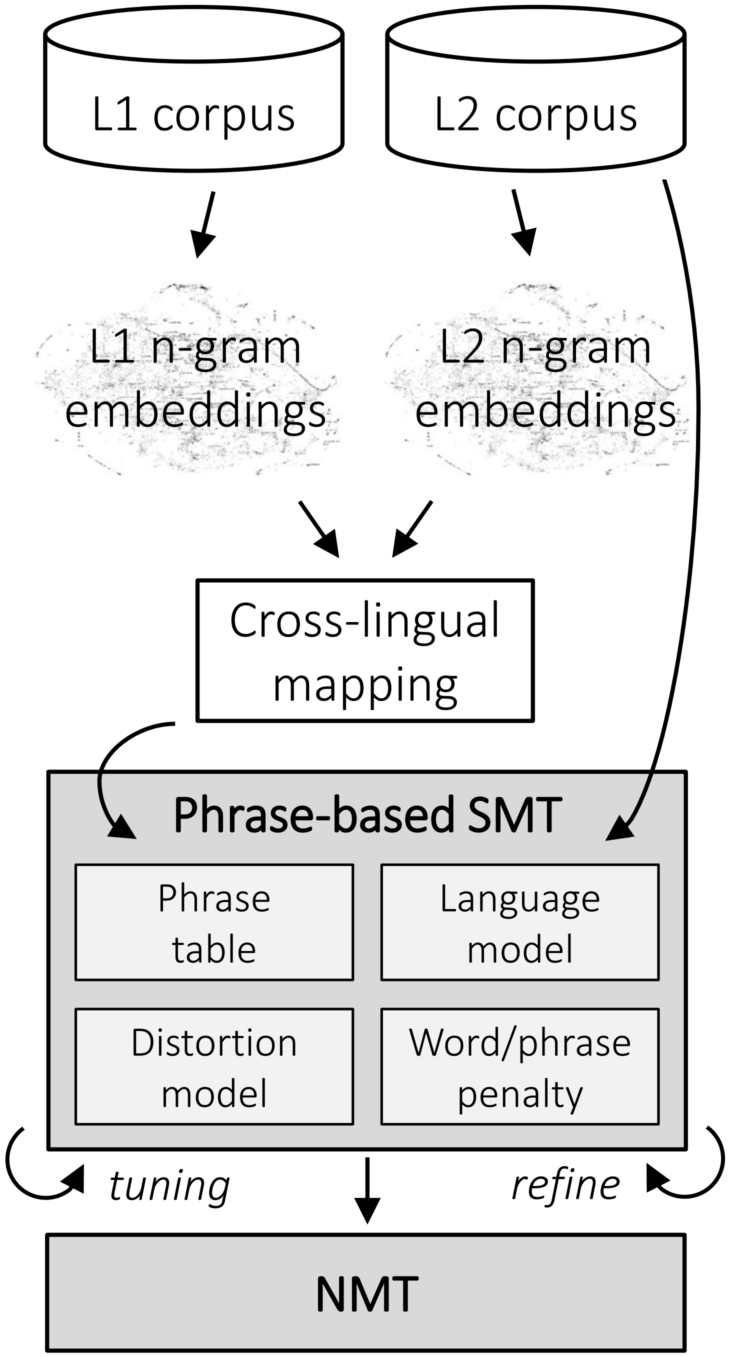
- Main recipe: iterative back-translation + smart initialization
  - Further explore the specifics? Start thinking beyond this paradigm?
- NMT will be part of the solution; SMT might or might not



What's next?

Some thoughts

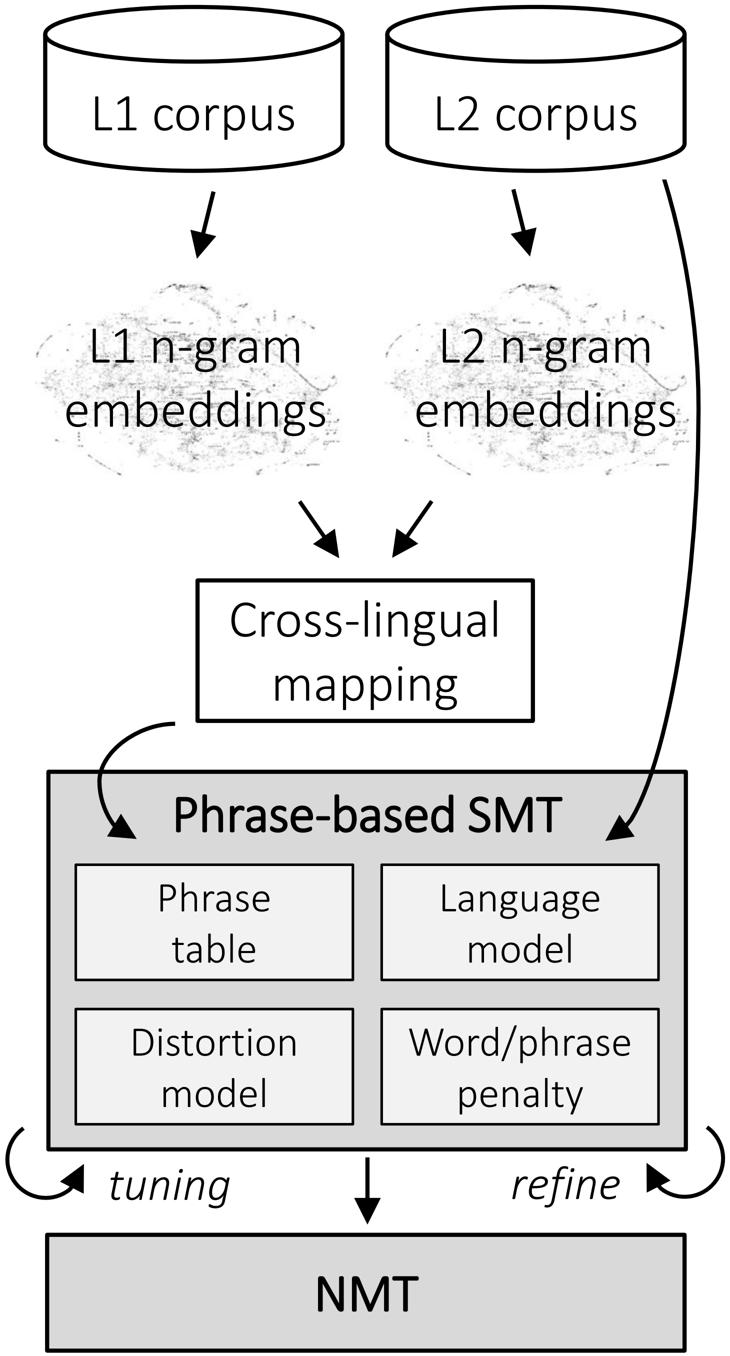
- Main recipe: iterative back-translation + smart initialization
  - Further explore the specifics? Start thinking beyond this paradigm?
- NMT will be part of the solution; SMT might or might not
- BERT is great, but I still think that starting from a lexical alignment and generalizing from that makes a lot of sense!



What's next?

Some thoughts

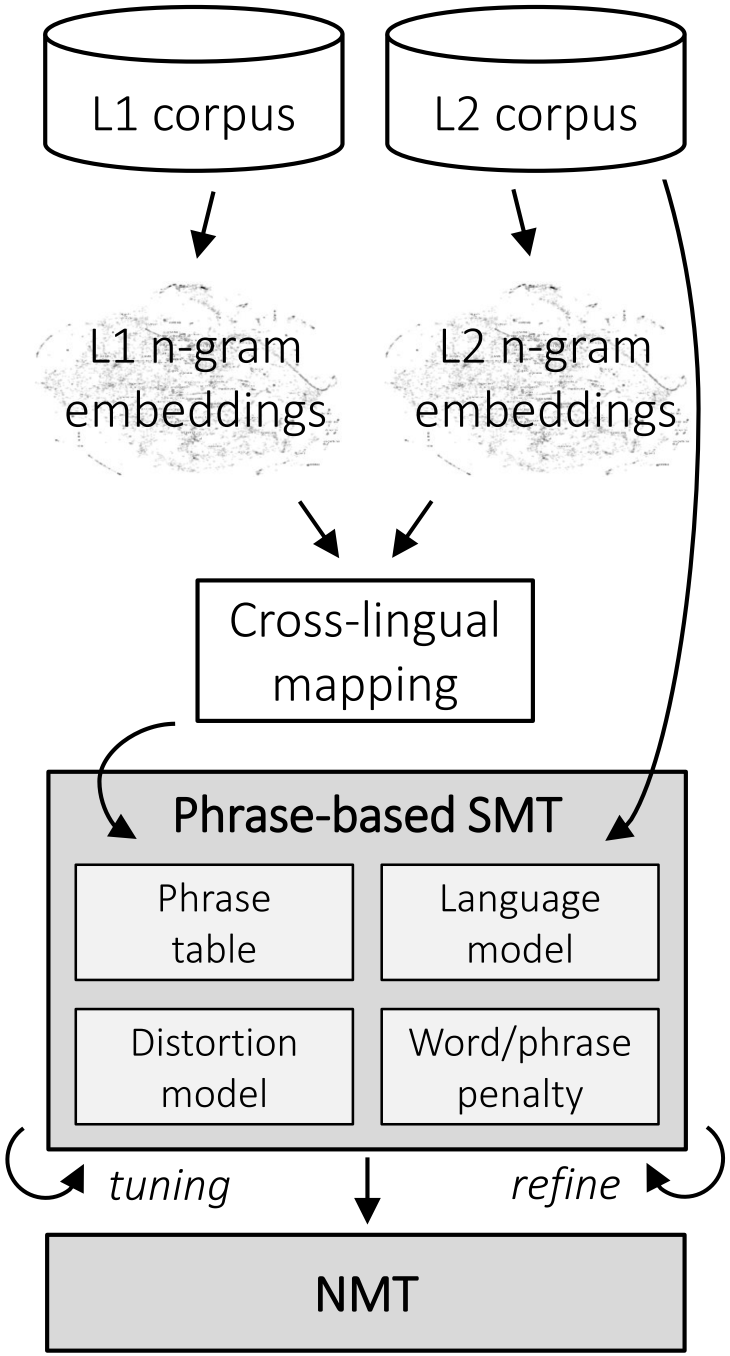
- Main recipe: iterative back-translation + smart initialization
  - Further explore the specifics? Start thinking beyond this paradigm?
- NMT will be part of the solution; SMT might or might not
- BERT is great, but I still think that starting from a lexical alignment and generalizing from that makes a lot of sense!
- More principled/interpretable approaches?



What's next?

Some thoughts

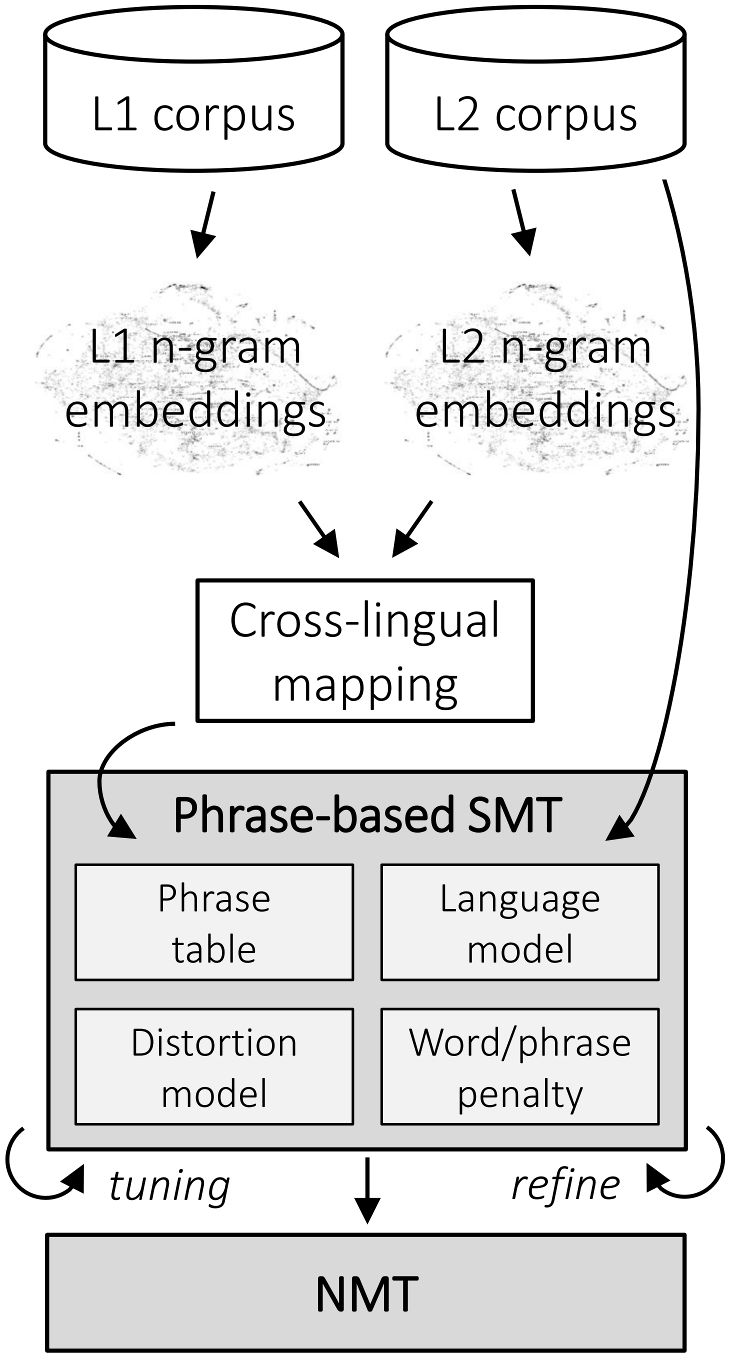
- Main recipe: iterative back-translation + smart initialization
  - Further explore the specifics? Start thinking beyond this paradigm?
- NMT will be part of the solution; SMT might or might not
- BERT is great, but I still think that starting from a lexical alignment and generalizing from that makes a lot of sense!
- More principled/interpretable approaches?
- Effect of linguistic distance, corpus size, noisy data...



What's next?

Some thoughts

- Main recipe: iterative back-translation + smart initialization
  - Further explore the specifics? Start thinking beyond this paradigm?
- NMT will be part of the solution; SMT might or might not
- BERT is great, but I still think that starting from a lexical alignment and generalizing from that makes a lot of sense!
- More principled/interpretable approaches?
- Effect of linguistic distance, corpus size, noisy data...
- We should think beyond immediate practical applications!



Thank you!

@artetxem

www.mikelartetxe.com